STATISTICS
B. A. E.
Lecture 2

FIELD OF STATISTICAL METHOD

REFERENCES: -

*Day - Chap. I. - Significance of Statistics.
- " II. - Variables and Statistical Units.

*Chaddock - " II. - Misuses of Statistical Data.

Lecture I traced briefly the historical development of statistics (i. e. statistical facts and their uses) from early history to the present. Today the quantity of statistical information is greater than ever before and the quality is steadily improving. During and since the World War, trade associations, banks, various State and Federal agencies, have begun the collection of a great variety of statistical facts needed as the basis of guidance for our highly complex industrial and agricultural civilization. All modern business on an extensive scale is based upon statistics - internal and external. The internal is usually based upon complete records, whereas external is more often based upon a sample of the universe.

1 - "Statistics may be said to consist of those methods of investigation peculiar to the collection, tabulation, and analysis of quantitative data". - Day.

2 - "Statistical method is the art of extracting the significant truths concealed in masses of numerical facts". - Jerome.
Statistical method consists of four processes: (a) Collecting; (b) summarizing; (c) analyzing; (d) presenting of statistical data.

(a) Collecting of statistical facts: Primary statistics are those collected by the person making the study, and carried on by questionnaire, correspondence, personal investigation, etc. Secondary statistics are those collected by other agencies and for other or more general purposes, - census data, market prices, car-lot shipments, etc.

(b) Summarizing of statistical facts includes sorting, tabulating, editing, classifying, and completing the collection so that a satisfactory count or sample is obtained.

(c) Analysis is the major part of statistical method. It consists of getting the "story" out of the data collected, includes frequency distribution, averages, dispersion, correlation, making of index numbers, time series analysis, etc.

(d) Presentation is arranging the data in such a way that the ordinary reader can understand the truths resulting from the analysis; includes table construction, making of graphs and charts, and explaining the significance of statistical measures obtained.

(Over)

3 - Personal requisites for statistical study and analysis:

(a) Intellectual integrity. Data must not be "doctored", processes inaccurate, or the facts misleading. Conclusions must be based upon facts.

(b) Constructive fertile imagination. Statistics is a search for truth, for causes or relationships. Details of problem must not overwhelm.

(c) Capacity for careful, painstaking work. Must not jump at conclusions before facts warrant it.

(d) Must know subject matter. A statistician is seldom only a statistician; usually an economist, sociologist educator, or business man, using statistics merely as a tool.

(e) Must be trained in proper statistical method. Proper and improper uses of statistics: Facts must be strictly comparable before accurate conclusions can be drawn. Statistical facts used largely for comparisons; if figures are not comparable, significance of the study is lost. Percentages must be used with caution, especially when applied to small numbers.

Variation is one of the most universal aspects of experience. Whenever differences find expression in quantitative data they may be made the subject of statistical inquiry.

Most of the material dealt with by statistical method in economics is not susceptible to laboratory experiment. Therefore, analysis can be made only by the careful breaking down of observed differences into their several elements, or of separation of items so as to render differences in particular groups of a more simple character.

A variable is anything which exhibits differences of magnitude or of number. The value of some variables are arrived at by mere tabulation of individual cases, some by specific measurements, others through computations, and still others by means of relatives or ratios.

Variables may be divided into continuous and discrete forms of variation. Continuous variables may assume any value whatsoever within the limits of their range, whereas a discrete variable can assume only certain positions on the scale.

Statistical Units. The unit is the individual case which is being considered in the study. It may be (1) a natural object; (2) concrete produced thing; or (3) institutional objects. For any particular study care must be taken to see that uniformity of definition of units is used throughout the study. Difficulties arise rapidly when explicit qualifications are introduced.

Variables may appear in a single object at different times, (Historical). There is also the variable element in a single object at different places, (Geographical). Variable may be expressed in repeated measurements of the same object (more common in natural science). Variation appears in the differing magnitudes of a single observed element in a number of related arguments (quantative). And objects will vary when classified according to given attributes (qualitative).

STATISTICS
B. A. E.
LECTURE 3.

THE USE OF SECONDARY STATISTICAL DATA

References: Jerome - Chap. 17 - Periodic Secondary Sources.

*Day - Appendix B - The Preliminary Examination of Secondary Data.

*Chaddock - pp. 392-395.

U.S.D.A. Bulletin 1480, pp. 54-64.

Original sources. Data gathered first-hand:

Government: Census, Crop Estimating Board Reports, Monthly Labor Review, Congressional Investigations.

Private: Commercial and Financial Chronicle, Annalist.

Semi-private: National Board of Economic Research, American Medical Association.

Compiled sources. Data transcribed from original sources.

World Almanac, Statistical Abstract, National Automobile Chamber of Commerce.

Data are often gathered for a special purpose and then applied to other studies where they do not fit at all. Data which are not really comparable are sometimes assembled in compilations and treated as if they were comparable, without adequate explanation of their limitations. Notes or bibliography should refer to basic sources.

Guiding principles in the use of secondary data: "It is never safe to accept published quantitative data at their face value". - Chaddock. The consumer of published facts must seek to find out for himself how they are collected and what they really mean.

"A first-class report of an original inquiry should furnish a copy of the schedule used, together with a careful explanation of the units enumerated, measured, or estimated. The method of selecting the samples and the procedure in collecting the data should be explained". - Chaddock. Then we can see if the information is representative, unbiased and reliable and whether it is adequate for the purpose of the problem. Where comparisons have been made, are they valued? Is the classification too crude for the purpose in view?

Facts are frequently used for a different purpose than that for which they were collected, (see pages 54-64 of U.S.D.A. Bulletin 1480, on utilization of farm price data).

Tests to be applied to secondary statistical data:

1 - Are the data applicable to the problems being considered?
2 - Are the data inclusive or complete for the "universe of inquiry"?

Accident risks basic data should include all accidents occurring, the place where and the condition under which they happen, and the extent of disability should be definitely known.

Checks on completeness very difficult to make. If data are available for small sub-groups such as county or preferably township and the enumeration has been made before as is the case with the Federal Agricultural Census, comparisons can be made between two successive enumerations.

3 - Are the data the result of sample inquiries, or estimates based on sample data? Then follows all the tests for sampling, representativeness, bias, fluctuation of sampling, etc., which will take up a large part of the first semester.

4 - How are the units of measurement defined? Are they simple or composite? Many errors introduced by the definition of the unit.

When ratios (relative numbers, or coefficients) are used, the denominators used should be scrutinized carefully. Effects should be related to the specific causes producing them.

5 - A fifth consideration is that the use of data is conditional also upon the (a) accuracy with which reported, (b) consistency of unit definitions, and (c) accuracy of determination.

(a) Accuracy with which data are reported and collected depends on the character of the informant, type of questions, asked, the care used in answering and recording them. Avoid difficult and unfamiliar or questions that might arouse distrust or suspicion, as replies are likely to be too brief or evasive. (Ex. Amount of personal debt). Age may be accurately known but falsely reported. "Seldom necessary to check numerical computations of reputable statistical publications; it is always necessary to satisfy oneself of the character of the primary material which is the basis for secondary tables." - Secrist.

(b) Accuracy of data is determined by the consistency with which definitions of units are used in the subject of the report, such as causes of death, uniform classification in all registration areas of the United States, standardized nomenclature with statistics of occupation, definition of capital. Definition of a farm and plantation. Production of milk for the year by enumeration.

(c) Accuracy of determination. Completeness may be impossible, as with estimates of total amount horse-power in our water resources, standing timber in United States. Use all estimates of this kind with caution.

Not all phenomena allow statistical measurement.

Degree of insanity not measured by number of insane patients in asylums.

Respect for higher education not measured by number of students in higher institutions of learning.

6 - Comparability of conditions which data describe is essential. To make price comparisons over the war period, war time prices are "deflated" to make them comparable with prices in more normal periods. This item has already been emphasized.

Statistical data may differ greatly in composition at different times because of changes in schedules used, editing, tabulating procedure.

STATISTICS
B. A. E.
LECTURE 4.

GATHERING OF PRIMARY DATA

References: *Chaddock - pp. 371-391.

*Jerome - Chapter 16.

Research Method and Procedure in Agricultural Economics,
Vol. I, pp. 21-31, 58-62, 104-126.

Planning the Investigation

When planning an investigation, one must keep in mind - (1) the relative importance of the several lines of investigation to which he might conceivably direct his attention; (2) the timeliness of the project; (3) available resources, - availability of cooperative resources; (4) the use of the product.

The planning of the investigation is the most important part of statistical procedure. It involves defining the purpose of the investigation, examination of available literature, choosing the universe, determining the scope of the inquiry, determining the method of collection that will give best results for money spent, selection of working force, arranging questions so as to avoid misunderstanding and to get all the data wanted in correct form. This is especially important in questionnaire method.

Arranging forms and instructions for use in field

Questions should be arranged so least personal come first, should be in logical order, should be so planned that a word or figure will answer, should be put on good quality paper of convenient size for filing and mailing. Questionnaires should be accompanied by letter explaining reason for investigation. Care should be used to see that no unnecessary questions are asked, no questions asked for which answers may be obtained elsewhere, and that schedules, etc., are never too long. Instructions should explain in detail just what is wanted, should define any units mentioned, should explain kind of sample wanted, type of people to collect from, and exactness expected in answers given as well as any checks that may be used on answers.

Methods of gathering data

1 - The questionnaire is a list of questions mailed out to persons, usually unknown and not expecting them.
2 - The correspondence method of investigation includes regular reporters who send in information at designated periods.
3 - The schedule method of inquiry is carried on by men in the field who ask the questions and fill in records.
4 - The personal investigation method is gathering data by personal contact with the facts.

Advantages and disadvantages of different methods

Questionnaire. - Least expensive, quickest, least satisfactory, smallest returns, must be short. Answers are likely to be biased, as only those interested answer.

Correspondence. - Extensively used in getting out periodic reports. Inexpensive, quick, fairly good returns once list is established. Even though reports are biased, once amount of bias is determined fairly accurate results can be obtained.

Schedule. - Much more expensive, more time necessary, reliability of returns depends upon agents gathering them. Only method possible on extensive projects.

Personal Investigation. - Very expensive, slow, and limited to small areas. Used by sociologists and in cost of production or standards of living studies.

Units and Measures. - Each unit must be rigidly defined: What is a family? a dairy cow? a farm? Should use units and measures in common use and generally understood. Much dispute over units. All doubtful or borderline cases must be anticipated and definition so formed as to avoid misunderstanding in any case. The comparability of statistical facts should always be kept in mind when working on definitions.

Statistical Universe. - A statistical universe is the area being studied; may be a factory, state, country, or world. After defining universe, be sure it is fairly represented. When using secondary data, the universe is determined by the area covered in the original study. Limits use of secondary data.

Testing out of questionnaire or schedule. - A small number should be tried out first to see that questions are clearly understood and can be answered correctly, and that all the data necessary for the study are obtained by the questions asked.

Sending out questionnaires. - Questionnaires should be accompanied by a self-addressed, stamped envelope to fit questionnaire. Should be accompanied by letter of explanation and instructions.

Editing of schedules and questionnaire. - Schedules should be checked for accuracy and completeness in the field by agents gathering data. Questionnaires are checked in office for completeness and correctness. Letters are sent to obtain corrections for errors and omissions. In editing, four points must be kept in mind - (1) accuracy; (2) consistency; (3) uniformity of expression; and (4) completeness.

After data are carefully edited, they are ready for summarizing and tabulating, but before going into that, the problem of sampling must be considered to see that all parts of universe being studied are fairly represented and that sample is actually representative of existing conditions in the universe.

STATISTICS
B. A. E.
LECTURE 5

SAMPLING

REFERENCES: *Jerome - Chap. II. - The Sampling Process.
Yule - Chap. XIII and XIV. - Sampling of Attributes.
"Research Method and Procedure in Agricultural Economics", pp.49-57.

Sampling is almost as important in the collection of statistical data as the planning of the investigation; in fact, it is part of the planning, as a perfect schedule will not give an accurate picture of universe unless the sample is representative.

Sampling is used in every statistical study made where the universe is too large for a complete count. It is the basis of all estimates, and the accuracy of estimates depend upon the accuracy of the samples from which they are made.

Objective of sampling

1 - To get an accurate description of conditions existing in a given universe of inquiry. Ideally the sample should be a "miniature" of the universe. Complete enumeration expensive, takes time to complete and is in many cases impossible to obtain.

2 - To obtain a measure of change in conditions rather than absolute conditions. The absolute level obtained in a sample may be too high, but the change from year to year may represent the change taking place in the universe of inquiry. Obviously, the "miniature" of the whole taken from time to time would reflect accurately the changes occurring in the whole. Ex. - The ratio of crop to crop from a complete census for two years crop, crop = x, would be identical with the ratio of the two ratios of crop to land in farm or the "ratio relative". Constant bias tends to be eliminated when data from two samples are used relatively.

3 - To obtain evidence of relationships between attributes or variables and to measure the extent of this relationship.

4 - To list the conditions that exist in a region. Different varieties of corn grown in a region.

5 - All of the above four objectives for cross-section analysis apply to "time series" analysis, but not in the same way. All we have is a sample, out of the infinite. Conclusions apply only to the period studied.

Laws of sampling

1 - The law of statistical regularity is that a moderately large number of items chosen at random from among a very large group are almost sure on the average to have the characteristics of the larger group. - Jerome.

(Over)

2 - The permanence of small numbers. "If among a great number of items there are a few which present some particular feature, it is a matter of common experience that this small number is seldom much exceeded and seldom entirely vanishes. - Bowley.

3 - The inertia of large numbers. - In the absence of any cause tending to bring a material change from one period to another, in most classes of phenomena, when one part of a large group is varying in one direction, the probabilities are that another equal part of the same group is varying in the opposite direction; hence, the total change will be slight. - King.

Methods of selecting a representative sample.

1 - Random selection - Usually known as "simple sampling". A sample found in such a way that everyone of the individuals in the universe of inquiry has the same chance of being selected in the sample, and that the selection of a particular individual does not influence the chance of selecting some other individual, "Selection in the basis of a lottery".

2 - Selection at regular intervals - Such as taking every tenth name in city directory or wages of every tenth worker in factory.

3 - Stratified random selection - The universe is subdivided into districts, geographically as crop reporting districts, - or on the basis of some variable as size of farm, tenancy, nativity of farmers, etc., - and a number of observations taken in each district. The districts may be made equal in size, or the returns may be weighted on the basis of the importance of the districts. Weighting may be automatic when the number taken is proportional to the importance of the district, or definite when an actual system of weighting is used.

4 - Purposive selection - Typical areas in the region to be studied are selected and all the units in that area are visited except the ones ruled out as outside the universe because abnormal. When a sample is secured by the "purposive method", groups of observations are deliberately selected by the statistician, the principle of randomness being entirely disregarded. The judgment of the statistician is substituted for impartial chance or the mechanical principle in the selection of the sample. It is assumed that the sample will have the same characteristics as the whole.

Upon what does the statistician base his judgment in making the selection of these districts or groups of units for the sample? As far as possible certain criteria will be used which relate to the field of inquiry. These criteria are called "controls". Controls are known for both the sample and the universe of inquiry, and are compared with the unknown quantities which are the subject of investigation. The selection of districts is so made that the aggregate of the districts gives the same results as the universe in respect to these control factors.

The size of the sample - The sample should be large enough to give a complete picture of the universe for the attributes being studied; therefore, the commonness of the attributes is an important factor in determining size of sample. Other factors determining the size of the sample are (1) accuracy required; (2) uniformity of universe; (3) details required in project; (4) expected return.

Test for accuracy - Take several samples, each consisting of about the same number of cases. Then if the results obtained do not show a reasonably close similarity, increase the size of the samples until the successive samples do evidence substantial similarity.

Increasing the accuracy of a sample - The precision of a sample increases with the size of the sample, varying directly with the square root of the number of observations in the sample. Thus to double the precision of a sample, it is necessary to quadruple the size of the sample; to treble the precision, it is necessary to increase the size of the sample nine times.

STATISTICS
B. A. E.
LECTURE 6.

The Theory of Probability and Errors in Sampling

References: *Jerome - Chap. X., The Theory of Probability and Error.

*Chaddock - 208-246.

The probability of the occurrence of an event may be defined as the expected relative frequency of the event in an infinite number of observations or trials. - Jerome.

Repeated observations of chance data in which there is a probability of an event occurring or not occurring, or in the accuracy of measurements, have shown that the items tend to distribute themselves about the most probable result in a symmetrical of bell shaped form.

The shape of this curve can be shown by a mathematical equation. On the basis of this phenomena of statistical data to conform to this bell-shaped curve, statistical theorists have devised formulas for measuring the most probable value of any group of data, the error to be expected in the most probable value, the probability that any item will fall within a certain range about this most probable value, and the probability that another group of data will show the same characteristics as the first group. It is by means of these formula that checks are made upon the accuracy of the sample. It should be remembered that these formula are based upon the assumption that the data, when properly summarized, follow the normal or bell-shaped curve. To the extent that they do not, the accuracy of the tests are invalidated. A large part of the first semester will be spent in studying the method of determining these measures and showing their accuracy. However, before going into the methods of determining the accuracy of a sample, it is necessary to examine why errors occur and how to avoid or measure them.

Mistakes:

Mistakes or "busts" arise from carelessness or incompetency in transcribing and reading figures or from numerical mistakes in computation. Statistical work is worse than useless if there are mistakes in it. All computations must be checked. The only safe assumption is that every one makes mistakes and a system of checking must be worked out that insures accuracy.

Checking can be accomplished by several systems.

(1) Have the original computations verified by another computer and the corrections verified by the original worker.

(2) The work may be handled in duplicate or triplicate.

(3) "Check sums", a form of cross addition, may be used as in multiple correlation.

(Over)

Standard of accuracy:

In the physical sciences a high degree of accuracy is practical but in the field of social and economic phenomena attempts to obtain the greatest possible degree of accuracy are, frequently, a waste of time. Ex. The average age of three boys of 4, 7, and 9 years is $\frac{20}{3} = 6.666$ years. It is absurd to carry this decimal more than one or two places. Statistics are used primarily in a comparative sense. It is not necessary to have the value of exports determined to the last cent. For every statistical problem, there should be determined, in advance, a definite standard of accuracy for each item, a standard that is practical and attainable with the funds available. Ex. (1) An age census need not be to the nearest day or month. (2) The area of a large lake in square miles would be accurate enough for all purposes, but with a small ice pond square feet would be needed. Statistical Accuracy should be thought of in the relative sense rather than the absolute.

Errors Encountered in Sampling

There are three general types of error found in the data themselves. The first type might be called Preventable Errors, ones that can usually either be prevented by the form and clearness of the schedule or eliminated by judicious editing; second, errors due to fluctuations of sampling which are measured by the standard error or probable error of the mean or whatever unit is being used; and third, cumulative or biased errors, which are the result of either, conscious or unconscious, personal prejudice or state of mind. In addition to these three general classes of error found in the data themselves is the more generalized error of lack of representativeness or the selectivity of the sample as a whole, especially when voluntary reporters are the source of information. There is also danger of having a biased or prejudiced statistician handling the inquiry.

Preventable errors

There are errors encountered in the sample data that must either be prevented or eliminated so far as possible. Experience with a given form of inquiry will bring to light a great variety of these errors such as:-

(1) Misunderstanding of the question or definition, - Ex. - condition or quality figures reported when yield per acre is asked, careful preparation of the "schedule" or "questionnaire" and tryout of the inquiry in a limited way will prevent most of these errors, but in case they occur they should be edited out if they can be detected.

(2) Reporting for a different unit or a different set of facts than those asked for because of local usage of terms. Ex. When the price of potatoes is asked per bushel a number of reports will come in on the price per 100 pounds, from those areas where potatoes are sold by the 100 pounds, rather than by the bushel. The yield of corn may be reported in three different units, - (a) per standard bushel equivalent to 56 pounds of shelled corn in the Corn Belt; (b) per bushel basket of ears in New York, Pennsylvania

and New England, and (c) per barrel of five standard bushels in parts of Maryland. The remedy is to ask for the price or yield in both or all units of measure, side by side on the same schedule.

(3) Confusing different years - reporting 1927 taxes in place of 1926. Preventable in most cases by proper construction of schedule.

(4) Incomplete or partial reply to questions. This difficulty arises when general questions are asked such as total crop acres, number of cattle, or hogs, receipts, expenses for labor, etc. If the component parts are asked separately and their totals combined, the results will be more complete. Ask in units in which the farmer thinks.

(5) Asking questions concerning facts for which the informer has no adequate basis of estimating, as total amount of milk produced during the previous year. Better repeat the inquiry periodically and ask for the production the previous day. Good common sense will prevent these questions being retained on the schedule. Unless you can ask the informer questions relating to his own experience and preferably immediate past or present, it is a waste of time and funds to conduct either a sampling inquiry or a complete enumeration.

Fluctuations of sampling

Errors of observation: Measurements of the same object repeated with the greatest care do not yield the same result. An engineer or astronomer measures a given distance several times knowing that the most probable value may be obtained from a number of observations by averaging the results. "Error in observation is the difference between the result of an observation and the true value of the quantity measured:" (x) The most probable value may be obtained by averaging the observations, providing the errors (or differences) of the separate observations are accidental and tend to balance each other, i. e. compensating.

Errors of observation are common to all scientific measurement. They occur in all statistical data whether it be a complete enumeration or registration or a sample from a universe of inquiry. Errors of observation are probably of much greater magnitude with social science data than with so called exact science data. It is difficult for a farmer to accurately estimate exactly how many acres of corn or wheat he harvested. The grain drill is about the only measure of acreage many farmers have available. Established fields generally acquire a specified area that may be handed down from father to son. Mere lack of exact knowledge on the part of the informer need not destroy the results of the inquiry. The estimate made too high by one may be compensated by that of another made too low and the average from a large number may represent the true value very closely.

Errors of observations must be given a broader interpretation in sampling analysis than when used in the field of astronomy where the concept was originally developed. It is not one single amount, distance or height that is being reported upon by the several observers. The purpose of the sampling process

is to secure an average, proportion or some other unit of measure from a sample that will coincide exactly with the average of the universe of inquiry.

The greater the variation in the universe of inquiry, the larger must be the sample if the resulting average is to have the desired "precision". The greater this variation the less likely is one sample drawn from this universe to be identical with another drawn at the same time and under the same conditions, or the greater will be the influence of the fluctuation of sampling on the average obtained from the sample.

Biased errors

Bias in its several phases is a form of error found in sample data as well as in complete enumerations and registrations. Biased errors differ from the errors measured by the probable error of the mean, - the fluctuation of sampling - in that biased errors are cumulative rather than compensating. They are errors that are "constant, presistent and biased". A very short person may read the thermometer hanging on the wall in such a way as to increase the reading every time an observation is made. No matter how many observations are made the average will never approximate the most probable degree of temperature. It is like using a short yardstick to measure a room. The prejudices or personal equation of the informer may influence him to observe only the phenomenon which support his news. This personal bias may be conscious or unconscious but the error becomes cumulative when any appreciable proportion of the observations are so affected.

Another important distinction between biased errors and the influence of fluctuation of sampling is the ease with which they are measured and their probable influence on the reliability of the average determined. The influence of the fluctuations of sampling can be measured quite satisfactorily from a study and analysis of the sample itself, unless the sample is too small; but biased errors can only be measured when the average for the universe is available as a check on the accuracy of the average of the sample after allowance for probable error has been made.

Bias caused by the personal equation:- The most difficult form of biased error to overcome or to make allowance for in making estimates for a universe of inquiry from a sample drawn from that universe is error, intentional or unintentional, that is the result of the prejudices or the personal equation of the observer or informer.

(1) Unintentional psychological bias;- tendency to exaggerate that which is the center of attention. For example, the county agents and bankers in a cut-over region where a land clearing campaign was underway, reported almost as much land cleared in two years as the census showed for five." (z page 48)

In years of propaganda of any kind concerning acreage changes or kinds of crops to grow there is always a distortion of the acreage samples in the same direction as the propaganda suggests:

(2) Memory bias: When reporters are asked to give estimates of the acreage of crops harvested on their farms last year along with the acreage for harvest this year we find that when a sufficient number of these reports are

compared with what was actually reported "currently" last year that the acreage of pasture and more important feed crops such as corn, oats, and hay check closely thereby indicating that memory bias was largely compensating and should therefore be classed as errors of observation. But with the minor crops the reporters seem to forget some of the acreages and the figures taken historically may under-estimate the acreage of these minor crops from 5 to 25 per cent or possibly more. This is memory bias and must be guarded against. With livestock the reporter tends to forget the calves and other young stock rather than the adult animals. This type of error is unintentional but cumulative and no increase in size of sample will overcome it. It is really a form of unintentional psychological bias.

(3) Wilful bias: There are sometimes intentional under or over statements on the part of the observer or informer and this may be called wilful bias. The classical example is the marked tendency for women to understate their age on an inquiry or to an enumerator. There is a marked tendency to underestimate the current year's acreage of cash crops on the part of farmers reporting on acreage schedules. The same error exists in reporting the yield per acre or production of an important cash crop such as cotton, until after the crop has left the grower's hands. Wilful bias is undoubtedly prompted by motives of self interest regarding the effect of supply estimates on prices that will be received for the crop, and the tendency is to be ultra-conservative in reporting supply factors to the agency that makes the official government forecasts and estimates.

Bias caused by selectivity of the sample

The sample as a whole may be "biased" by the fact that it is not fully representative of the universe of inquiry. The returns from a mailed questionnaire on acreage of crops may come from farmers having larger farms than the average. The picture of the proportion of the farm acreage in different crops would be distorted as large farms might be expected to have a somewhat different utilization of land than small farms. The same difficulty arose when a government commission collected costs of growing peanuts from farmers who kept some sort of accounts. It was found that the yield per acre on the farms studied were much higher than the average for the region and consequently their costs per bushel were much lower. The study was not representative of the universe of inquiry.

Bias caused by the lack of representativeness or selectivity of the sample, can make a sample which is excellent in all other particulars worthless, especially when the objective is a "miniature" of the universe of inquiry. (Obj. 1). When the sample is to be used only to measure changes or the trend of conditions (Obj. 2) the presence of a constant degree of "Selectivity Bias" can be overcome by using successive samples in a relative sense. Ex. Farm prices collected each month are used as the basis for constructing an index number of farm prices. The great difficulty with bias which is due to selectivity of the sample is in knowing whether or not it is really constant from one inquiry to the next and over a period of time. When objective 3 is paramount "To obtain evidence of relationships between attributes or variables and to measure the extent of this relationship," even substantial longtime changes in the degree of bias may be taken care of by multiple correlation technique. Ex. - The condition of a crop in per cent of "normal" may have a large amount

of either constant bias, or slowly changing bias without impairing its usefulness as basis of forecasting the probable yield per acre of the crop.

With sample inquiries that apply to acreages of crops and numbers of livestock etc., where a census has been made within a year or two previous, a rough test of the representativeness of the sample is to determine, first, the percentage that all land in farms is of the total reported by the census and then see how closely the other factors approach this same decimal or fraction of the whole. Ex. - If the sample contained 1/20 of the land in farms does it also contain 1/20 of the work animals, the pasture, the number of tenants, etc.?

Statistician's bias

Sample data may be so "edited" and otherwise handled by an unskilled or prejudiced statistician in such a way as to distort the picture. All the very high yields per acre may be edited out as not probable, when such a procedure would not be justified by the facts of the situation. It is a dangerous practice to attempt to complete a schedule that is not entirely filled out by the reporter.

Characteristics of a Sample

In attempting to evaluate a sample, whether it is one that you have collected yourself, or one that someone else has collected, there are about five pertinent questions that might be asked;

1. Is the sample representative of the universe of inquiry? If the observations in the sample are obtained from voluntary reporters the sample is likely to be selective of the more intelligent persons who are willing to take the trouble to reply to an inquiry. Dividing a State up into districts and weighting each district helps to improve representativeness. Breaking up the acreage sample into size of farm groups and weighting each group helps to insure representativeness.

2. Is the sample large enough to give precision to the average, or other measure used? Increasing the size of the sample reduces the influence of the fluctuation of sampling as well as the effect of errors of observation (non-cumulative or compensating).

3. Is there Bias in the individual observations?

   (a) Intentional?
   (b) Unintentional?
   Check data necessary if bias is to be measured.

4. Are there any Preventable Errors in the sample? Is the questionnaire adequate and also understandable? In how many ways can the question be incorrectly answered? To what extent is the editing practical?

5. Is the agency or statistician making the study likely to be biased or prejudiced in any way?

Statistics
B.A.L.
Lecture 7

CLASSIFICATION AND TABULATION

References:

Jerome - Chap. III
Secrist - Chap. V
Handbook of Instructions for Standardizing Tabular Forms and Methods - U. S. D. A., Division of Statistical and Historical Research -

Classification is the arrangement of items into groups according to common characteristics. Classification must be exclusive, i.e., there must be no overlapping of subclasses.

Methods of Classification - historical, geographic, qualitative and quantitative. By methods of cross-classification two or more of these characteristics may be shown at the same time. Data may also be classified according to magnitude or size by use of a frequency table.

Mechanics of classification

In a well planned statistical investigation the scheme of classification is determined upon before the work of gathering data is commenced. The data may be tabulated by hand or by machine. In large organizations tabulating is done by machines by use of punched cards. When tabulation is done by hand prepared tabulation or tally sheets facilitate the work greatly and cut down the possibility of error.

Statistical tables

After tabulating, sorting and counting are completed, the next step is to record the data in tabular form. Tabular forms may vary according to the particular purposes for which they are to be used. There are, however, certain principles of technique which are of general application and should be observed in the construction of statistical tables.

We collect data, summarize them, and draw certain conclusions therefrom. The role of the table is to present these data in such form that the story we wish to tell is set forth in definite, clear, understandable form.

The data themselves will determine largely what kind of table we should use. It is generally wise to avoid highly complex tables. There is a limit to the amount of detail the eye can grasp. It is wise, therefore, where complex conditions are to be described to present the material in several tables rather than to present too much in one. Too great detail often suggests confusion and repels attention. Jerome says, "A good statistical table is not a mere careless grouping of columns and rows of figures; it is a triumph of ingenuity and technique, a masterpiece of economy of space combined with a maximum of clearly presented information".

Terminology of tabulation - "A table is the arrangement of figures in columns and rows so arranged that the significance of the figures is determined by two sets of phrases, captions at the head of the column and stub at the end of the rows"- Jerome. See table I.

The year 1921 is a stub and the figures opposite 103,748 - 29.6 - 3,068,569 are spoken of as a row. The word "Acreage" is a caption and the figures beneath are spoken of as a column. Tables vary considerably, but they are all made up of some combination of the four elements captions, stubs, rows and columns of figures.

The Title - The title is the sign post, it is to the statistical table what the heading is to the news item. Five principal requirements for the title are:

1. It must be concise. It should begin with the key word.

2. It should be catchy - attractive - draw attention.

3. Clearly understandable and impossible of double meaning.

4. It must be adequate, answer the questions - what, when and where.

Titles of tables in a series should be preceded by the word "Table" and the number, period and dash. Dates showing period covered should be placed at the end of the title,

ex. Table 4.- Corn: Acreage, yield per acre and value in the United States, 1918-1928

Ruling - The purpose of ruling is to bring out the relation of the several parts and at the same time increase the attractiveness. Some tables are unruled, but this is to be avoided in complicated tables.

1. Use of the open side or semi-box type is most common.

2. A total or an average should be set off by a single line above and where other data follow by a double line beneath. The double line may be omitted if the figures are in bold face type.

3. Main divisions should usually be set off by double lines.

Stubs and Captions - In complex Tables it is often a good plan to number both columns and lines.

The stub and captions should be brief and at the same time describe clearly the figures following. When the wording for one classification is much longer than another, it is usually a saving of space to place the longer on the stub.

The contents of the Table - In arranging the order of stubs and captions, it must first be decided the order of emphasis.

Totals are preferably placed at the top of the columns. This is done because many people are interested only in the totals and by placing them first it gives them the most conspicuous place in the Table. (However, the Bureau of Agricultural Economics usually places totals at the bottom or right of a Table.)

Arrangements for emphasis in the Table are classified as follows:

First - or primary comparisons - adjacent in the same column.

Second - or secondary comparison - adjacent in same row.

Third - alternating in same column.

Fourth - alternating in same row.

There are necessarily modifications to these rules. For example, if a comparison of crop yield for all States for five or six years is to be made, it would probably conserve space to place the States on the stub even though it is desired to place primary emphasis on the comparison between years.

Table 1.- Corn: Acreage, yield per acre and production 1/
in the United States, 1921-1925

Year	Acreage	Average yield per acre	Production
	1,000 acres	Bushels	1,000 bushels
1921	103,740	29.6	3,068,569
1922	102,846	28.3	2,906,020
1923	104,324	29.3	3,053,577
1924	100,863	22.9	2,309,414
1925	101,395	28.8	2,916,961

1/ Bushels of 56 pounds.

STATISTICS
B. A. E.
LECTURE 7

Classification of second hand fords at A. Bee Ford Agency, January 1, 1930, showing the type, model, year of car and kind of wheels

Type of car and year of make	Model of car and kind of wheels								
	Total cars			Wire wheels			Wooden wheels		
	Total	Model A	Model T	Total	Model A	Model T	Total	Model A	Model T
	No	No	No	No	No	No	No	No	No
Total	163	88	75	104	78	26	59	10	49
Roadster	25	16	9	16	15	1	9	1	8
Touring	13	5	8	9	5	4	4	0	4
Coupe	48	27	21	32	25	7	16	2	14
Tudor	44	22	22	28	21	7	16	1	15
Fordor	33	18	15	19	12	7	14	6	8
1928 model total	118	65	53	79	57	22	39	8	31
Roadster	18	13	5	13	12	1	5	1	4
Touring	9	4	5	7	4	3	2	0	2
Coupe	35	20	15	24	18	6	11	2	9
Tudor	34	18	16	22	17	5	12	1	11
Fordor	22	10	12	13	6	7	9	4	5
1929 model total	45	23	22	25	21	4	20	2	18
Roadster	7	3	4	3	3	0	4	0	4
Touring	4	2	2	2	1	1	2	0	2
Coupe	13	8	5	8	7	1	5	0	5
Tudor	10	6	4	6	4	2	4	0	4
Fordor	11	6	5	6	6	0	5	2	3

Statistics
B. A. E.
Lectures 8 and 9

METHODS OF GRAPHIC PRESENTATION

References:

Jerome - Chap. 4 - 7 - Graphic Presentation
Brinton - Graphic Method for Presenting Facts

Why use graphs

"The use of diagrams and graphs represents the effort of the statistician to utilize the attention-attracting power of visual presentation. By translating dry figures into graphic forms he hopes to command the interest of the reader and at the same time to give him a correct impression of the essential facts." - Jerome.

When to use graphs

Graphs are expensive and should be used only when they are necessary to attract attention to facts or whenever the real significance of the facts cannot be brought out without them. Graphs or pictures are commonly used in advertising to attract attention to significant points in development or services of an industry.

The principal points to keep in mind when using graphs are (1) be sure the graphs give a correct picture of the facts, (2) that the graphs are interesting and attractive, (3) that they show the facts in the correct manner, and (4) that they are verifiable by the reader.

Maps are used for showing geographical locations such as principal markets, soil regions, and the distribution of population, of percentage of indebtedness, amount of rainfall, or length of growing season. The common essentials of all maps are a title, date, legend and reference. Title should be brief and to the point. Legend should be clear and where classes are used they must be uniform in size. A map should be accompanied by a table showing more accurately the data being illustrated.

A few of the more common types of maps are listed below.

(1) Distribution or small dot chart maps. Dots are of uniform size with each dot representing a constant amount.

(2) Quarter dot maps - used for same purpose as small dot maps but where area is not as well known - gives less definite location than small dot map.

- Over -

(3) Large dot maps - where size of dot represents its importance - used to show location of principal markets, size of towns, etc. - should always be accompanied by a table showing basis for dots.

(4) Isogram or Contour maps. Lines are drawn through map dividing areas according to classification - classes must be equal - lines smoothed within limits of accuracy - and each line well labeled. Used for showing length of growing season or date of killing frost, etc., where all areas have equal importance.

(5) Shaded Isogram map - same as (2) except each different class is shaded or cross hatched. Used in showing amount of annual rainfall, average temperature, etc., where certain areas want to be given more importance.

(6) Cross hatch maps - where areas are shaded according to importance - their best use is to bring out differences in percentages, soil types, agricultural regions, etc. May be used to show business conditions, sales, areas, etc.

Bar Charts may be roughly divided into vertical bars (for comparison over period of time), horizontal bars (for comparison at a given time), histograms, component part bars, and percentage bars. Every bar chart should have a clear concise title, a number when more than one chart is shown, scale elements (scale points, guide lines and designation), key or legend data to verify chart, references, and a box enclosing the chart. Bars should be proportioned to sheet; should stand out, as the object of chart is to attract attention; and should not be too complicated as their ability to attract attention disappears rapidly as complications increase.

Picture bars - same as horizontal bar charts except pictures are used for bars. More likely to attract attention.

A pie diagram is a circle divided into sections to show distribution of income, costs of production, kinds of production, etc. Often used for percentages also. They are noted for attracting attention, but difficulty arises in comparing sectors. Two or more pie diagrams should never be used in making comparisons. Bar charts are much less confusing for this purpose.

Connecting Component bar charts are used to show percentage changes in each part over a period of time.

Organization charts are used in large firms to show the processes of production, responsibility of employees, or the division of the different branches of the plant into its component parts. Are a great aid to heads of extensive corporations.

Rules for bar charts (Brinton).

1. Avoid using areas or volumes when representing quantities. Never use different sized pictures. Presentations read from only one dimension are least likely to be misinterpreted.

2. The general arrangement of a chart should proceed from left to right.

3. Figures for the horizontal scale should always be placed at the bottom of the chart. If needed it may be placed at the top also.

4. Figures for vertical scale should always be placed at the left of a chart. If needed may be placed at the right also.

5. Whenever practicable, include in the chart the numerical data from which the chart is made. If not in the chart it should be shown in an accompanying table.

6. Make the title of a chart so complete and so clear that misinterpretation will be impossible.

Line graphs have less ability to attract attention than charts; but are far more easily made and show data more accurately. They can also be made to attract attention by shading the area underneath or between the lines. Line graphs are used for presenting historical series and can be made much more complicated than charts. Line graphs should be used only in a continuous series.

Rules for constructing line graphs

1. The time scale should be on the horizontal axis and the quantity scale on the vertical axis.

2. Curves drawn on arithmetically ruled paper should show the zero line whenever possible and if not a break should be made in the chart to show that part of the chart is missing.

3. The zero line should be a much broader line than the average coordinate lines.

4. When the scale of a curve refers to percentages the line 100 per cent should be a broad line the same as the zero line.

5. Make curves of much broader lines than the coordinate ruling so that curves can be clearly distinguished from the background.

6. Each curve should be a distinct type of line and a legend should identify the series that each line describes.

- Over-

7. Curves should also be accompanied by a table showing the data plotted or a reference should be given where data can be found.

The Ratio chart is a type of graphic presentation suitable for showing proportional changes, particularly when it is desired to indicate the absolute amounts at the same time. The vertical scale is a logarithmic scale while the horizontal scale is an arithmetic scale. The same results are obtained by plotting natural numbers on a ratio chart as would be obtained if the logarithms of the numbers were plotted upon a natural chart.

A ratio chart has no zero on its vertical scale as the ratio between zero and any number is infinite.

It is not absolutely necessary that one understand logarithms to construct and employ ratio charts. However, in as much as ratio ruling rests upon logarithms a brief review of the nature of logs may help to insure a clear grasp of the nature of a ratio chart.

A logarithm of a number is that power to which ten must be raised to obtain the given number. To multiply two numbers their logarithms are added, to divide they are subtracted. To square a number the logarithm is multiplied by two; to cube a number its logarithm is multiplied by three, etc. On a ratio chart (1) a geometric progression is a straight line (2) equal vertical rises represent equal proportional changes (3) equal slopes of two curves or of two segments of the same curve indicate equal rates of change.

Statistics
B. A. E.
Lecture 10.

The Frequency Distribution

References: Mills - Chap. 3 - Organization of Statistical Data.
Crum and Patten - Chap. 9 - Charting: Frequency Series.

One of the most common methods of arranging unorganized data is to classify them into well defined classes according to size. Such a series is called a frequency distribution.

The Array: Is the arrangement of figures in order of magnitude. It is not necessary to arrange data into an array before making a frequency table, but it is desirable where a detailed study of individual items is to be made.

The frequency table presents a summary of the original figures. This condensation results in some loss of detail, but makes it possible for the mind to more readily grasp the full significance of the data.

Structure of Table: (1) Should have a clear, concise, complete title. (2) Headings of columns and rows should be concise and unambiguous. (3) Variable quantities should increase from left to right and from top to bottom. (4) Unit of measurement should be clearly indicated. (5) Sources should be given in all cases and (6) Table should constitute a unit self-sufficient and self-explanatory.

The Class-interval: The smaller the class-interval the less the loss of detail but the table is more cumbersome. Unless the size of the class-interval is determined by the data, the smallest and largest item in the data should be found and the difference divided into from 12 to 20 classes in order to get the best sized table. Classes should be so arranged that there will be no material departure from an even distribution of cases within each class. An orderly and regular sequence of frequencies should be secured. Class intervals should be uniform throughout the table.

Location of class limits should be such that the midpoint is a whole number. Where data tend to group themselves around given values at regular intervals classes should be so arranged that these values are the midpoints of the class.

Definition of classes: Classes must be arranged so there is no uncertainty as to where any given case falls. This can be done by leaving a small interval between the limits of each class. There should be no determinate classes.

Graphic Presentation of Frequency Distributions: In constructing the graph frequencies are always arranged on vertical scale and class intervals on the horizontal scale.

A histogram is a vertical bar chart, each class representing a bar and bars are placed adjacent to each other.

Frequency polygons are formed by joining the midpoints of each vertical bar by straight lines. Where the data are taken from a sample the polygon should be joined to the base line at the midpoint of the class just below the lowest class and to the midpoint of the class just above the highest class for which data are available. When based upon a complete count the lines are joined at the lower limit of the smallest class and the upper limit of the largest class.

A frequency polygon is a picture of the distribution of the sample taken from the universe. When smoothing the polygon into a frequency curve the curve is made to give a picture of the distribution of all the data in the universe. Much care must be used in smoothing the curve, especially when the polygon is irregular to be sure the curve is drawn to represent an actual picture of the distribution. The extent to which the smooth curve can deviate from the polygon depends upon (1) the knowledge which the statistician has of the problem being studied (2) the size of the sample and (3) the tendency of the data to follow a definite type of distribution.

Cumulative arrangement of statistical data: Data may be arranged cumulatively in a frequency distribution, either in ascending or descending order. Cumulative distributions are commonly used inlength of life tables, or depreciation tables so that total frequencies at any given point can be readily determined.

When plotting a cumulative frequency it is called an ogive. The ogive is valuable in forecasting totals or estimating depreciation at any given point on the frequency scale. An ogive curve is constructed in the same way as a frequency curve except the cumulative frequencies are plotted.

Frequency distribution curves and ogives have many uses and should therefore be constructed with considerable care. They represent the universe and by interpolating from them it is possible to determine conditions in the universe at any point on the frequency scale. They give a picture of the universe so that an idea of the extent of skewness, the amount of dispersion and the most probable value can be readily seen at a glance. However, these things can only be seen approximately on the curves and more exact measures are needed for detailed study. These are obtained by the use of averages, measures of dispersion, kurtosis, and skewness.

Statistics
B. A. E.
Lecture 11.

AVERAGES

Reference:

Mills - Chapter IV.
Yule - Chapter VII - Averages.

A frequency distribution gives a picture of the data being studied and arranges the data in a manner which enables more comprehensive studies to be easily carried out, but it only enables qualitative or verbal comparisons to be made. The next step is to calculate the quantitative measures of the distribution so that more detailed comparisons may be made between two or more series. In a normal or moderately assymetrical distribution the average is the most valuable quantitative measure to be applied to the distribution.

An average (1) is a measure of central tendency; (2) is the most probable value of the distribution; (3) is a single value of the variable which is representative of all the values of the variable included in the series; (4) is a typical value of the variable.

The characteristics of a good average (Yule)

(1) An average should be rigidly defined; (2) should be based upon all observations made; (3) should not be of too abstract a mathematical character, that is, its general nature should be readily comprehensible; (4) should be calculated with reasonable ease and rapidity; (5) should be as little affected as possible by fluctuations of sampling; (6) should lend itself to algebraic treatment.

Types of average

1. The arithmetic mean is the sum of the observations in a series divided by the number of observations. It meets all the requirements of a good average. All items in the distribution have equal importance in determining the arithmetic mean.

2. The geometric mean is the Nth root of the product of the observations. It is calculated by using the logarithms of each observation, getting their sum; dividing it by the number of observations and converting the result back to natural numbers. It meets all the essentials of a good average, except essential (4). It gives more weight to the smaller numbers in a distribution and is therefore always smaller than the arithmetic mean.

3. The harmonic mean of a series of measures is the reciprocal of the arithmetic mean of the reciprocals of the individual measures. In calculating it the reciprocal of each number is determined, their sum found and divided by the number of observations. The result is then converted back to natural numbers by taking the reciprocal of it. It meets all the essentials of a good average except essential (4), but its use is limited. It must be used in averaging time rates and where extreme emphasis is to be placed upon the smaller items of a distribution. Its value is always smaller than the arithmetic or geometric mean.

4. The quadratic mean is the square root of the arithmetic mean of the squares of all the observations. It is used mostly in calculating the standard division of a series and tends to give undue weight to the extremely large items in the series. It is therefore larger than all other averages.

5. The mode is the value of the observation that occurs most often in a series. It is easily found in an array by inspection, is assumed to be the value of the highest point on a frequency curve and can be calculated by inspection from a frequency distribution. It does not meet essential (2) and may not meet essential (1) or (6). One of its chief values is its ease of calculation or approximation.

6. The median is the middle number of a series of observations and is calculated by inspection of an array or from a frequency series. In a small number of observations it is often used because, although all observations are influencial in determining its value their number rather than size is the important factor and an unusual item will not distort the value of the median.

In a normal distribution the value of the mode, median and the arithmetic mean will be the same, but in an assymetrical distribution which is skewed to the right the mode will be smallest, the median next and the arithmetic mean will be the largest of the three.

ILLUSTRATIONS AND LEGENDS

1. Hand in one or more bulletins, circulars, or similar publications, which contain an illustration and also a legend which you consider faulty, and give your reasons in detail on a sheet attached to the printed matter.

2. From the following data construct graphs that will show the desired relationships clearly and most appropriately.

A. One of the tomato diseases develops spots on the fruits. The spots enlarge more rapidly on ripe than on green fruit, as shown by experiments the results of which are here given. At the beginning of the experiment the tomatoes were separated into three lots -- ripe tomatoes, nearly ripe tomatoes, and green tomatoes. The tomatoes in each lot were separated further into lots on the basis of size of spots on the fruits. The diameters chosen for the classification were 0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, and 5.0 millimeters. At the end of the experiment it was found that the spots had increased in diameter as shown below:

Diameter of spots at beginning of experiment	Increase in diameter of spots on ripe fruits	Increase in diameter of spots on nearly ripe fruits	Increase in diameter of spots on green fruits
Mm.	Mm.	Mm.	Mm.
0.0	0.29	0.18	0.12
0.5	.27	.17	.11
1.0	.24	.15	.09
1.5	.22	.12	.08
2.0	.21	.11	.07
2.5	.19	.10	.06
3.0	.17	.09	.05
3.5	.16	.08	.04
4.0	.14	.06	.03
4.5	.13	.05	.02
5.0	.12	.04	.01

B. An experiment was conducted to determine the temperature of the bark on the north side of some western yellow pine trees as contrasted with the temperature of the bark on the south side of the same trees. The following data were obtained.

Time at which temperature was taken	Temperature on North side °F.	South side °F.
10 A.M.	43	50
11 A.M.	45	58
12 M.	57	66
1 P.M.	58	72
2 P.M.	59	75
3 P.M.	56	60
4 P.M.	48	52

3 . The octavo-size publications of the Department allow for a maximum size of 4 1/8 x 7½ inches for illustrations, including space for legends. Assume that one-half inch will be required for the legends in the following examples, whether the illustrations are placed narrow or broad. Mark each of the illustrations here given for reduction or enlargement so that one side will be maximum width or length and at least one half inch be left on the other side for the legend.

Problem: What will be the other dimension in each case ?

Illustration 1 - 5 x 7 inches

Illustration 2 - 8 x 16¾ inches

Illustration 3 - 2½ x 6 inches

Illustration 4 - 3¾ x 5½ inches

4. Edit the following legends according to the Department style.

Fig. 1: -- Plows used on the experimental Farm -- a: disk plow.

b.Sulky Plow; C, moldboard Plow.

Figure 2 -- : Roots of plants after they were exposed to the air:

Natural Size; A: Corn roots as they appeared 24 hours after exposure:

No root hairs developed on them. b.--wheat roots 2 days after exposure.

X 5.

Fig. 3. -- a: cross section of maple stem. B. -- longitudinal section of Maple Stem; c; Radial section of oak Stem.

Figure 4, A. -- The experimental plots as they appeared when the work was begun. Seventy plots were used. b: the wheat of five varieties grown on the plots. The photograph was taken just before harvest. Photographed by the Jones Photo Co.

The following 10 examples are submitted to test the student's knowledge and judgment of correct form in the preparation of tables. They are taken in whole or in part from Department publications or from manuscripts submitted for publication. In each example are common mistakes which the student is expected to correct. Corrections should be made on the copy, but if any table requires transposing new copy should be prepared. From the textual material the student will prepare a suitable table. The figures in all cases may be considered correct, so no time need be spent on verification.

Table 1. -- Vegetables Containing Largest Quantities of Minerals

Calcium	Phosphorus	Iron	Iodine
Turnip tops	Kidney beans, dry	Lentils, dry	Highest
Chard	Lentils, dry	Kidney beans, dry	in
Kidney beans, dry	Peas, dry	Lima beans, dry	leafy
Cauliflower	Lima beans, dry	Peas, dry	vegetables.
Dandelion	Lima beans, fresh	Spinach	Amounts
Lentils, dry	Peas, fresh	Dandelion	vary in
Endive	Brussels sprouts	Chard	vegetables
Peas, dry	Mushrooms	Lima beans, fresh	from
Celery	Corn	Watercress	different
Kohlrabi	Parsnips	Peas, fresh	localities.

Table 2. -- Width of rings in Western Larch boards

Rings per inch	Grades				
	C and Better	D Select	No 2 and Better	No. 3 Common	All Grades
Number	Percent of total number of pieces				
4	---	---	---	0.2	.1
5-6	---	---	0.5	1.5	.6
7-8	---	---	2.9	5.6	2.3
9-12	0.4	1.5	4.8	10.4	4.8

Table 3. -- Length of term of farm-mortgage loans: Percentage distribution of holdings of principle lending agencies

Agency	Average term	Term					
		1 year	2 to 4 years	5 years	10 years	11 to 30 years	over 30 years
	years	Per cent	Percent	Per cent	Per cent	Per cent	Per cent
Insurance companies	5.6	4.4	13.3	64.8	14.6	2.5	0.4
Federal land banks	33.0	---	---	---	---	---	100.0
Joint-stock land banks	33.0	---	---	---	---	---	100.0
Commercial banks	2.6	52.1	19.9	26.7	0.7	0.6	---
Mortgage companies	6.2	0.3	2.8	74.5	20.6	1.8	---
Other sources	4.7	20.1	13.5	53.6	11.1	1.7	---
All agencies	8.5	17.5	11.7	46.5	9.6	1.5	13.2

Table 4. -- Estimated vegetative composition palatability and grazing capacity of a representative mixed-grass type a/

Species	Composition	Palatability	Palatability factor
	Per cent	Per cent	Grams
Hilaria mutica	37	60	.102
Sporobulus auriculatus	25	20	.024
Boutelona barbata	16	10	.001
Lesquerella fendleri	14	20	.002
Ephedra trifuria	10	20	.004
Totals	100	--	.135

a/ Reconnaissance plant density 0.4
Forage factor 0.138
Approximate number of surface acres required to support a cow for 1 month 5.80

Table 5. -- Repellent effects upon Japanese beetles of various toxic and nontoxic materials

1/ Plus sign indicates attraction

Date	Duration of experiment	Materials used per 50 gallons of water							
		Lead arsenate 3 pounds		Slaked Lime 4 pounds		Chalk 3 lbs.		Paris green 9 lbs.	
		Beetles on tree	Repellency	Beetles on tree	Repellency	Beetles on tree	Repellency	Beetles on tree	Repellency
	Days	Number	Percent	Number	Percent	Number	Percent	Number	Per cent
July 10	10	1,931	90.4	12,166	39.7	---	---	---	---
July 15	12	1,008	91.9	1,762	85.8	---	---	---	---
July 21	6	2,247	66.1	---	---	---	---	---	---
July 30	5	---	---	---	---	291	69.7	1,722	+79.4 1/
Aug. 2	3	8	94.8	33	78.4	---	---	1,019	+556.0
Aug. 6	2	25	86.0	---	---	---	---	386	+115.6

Table 6. -- Annual cost per acre of operating a nut orchard and yield required to cover costs

Item		Quantity per acre	Cost per acre
			Dollars
Labor and power prior to harvest			
Man labor	hours	26.0	6.82
Horse work	do	8.5	1.28
Tractor work	do	2.3	2.87
	Total	--	10.97
Materials			
Fertilizer at $37.50 per ton	pounds	570	10.69
Winter cover crop, vetch at 12¢ per lb.	"	20	.84
Miscel.		--	60
	Total	--	12.13
Other costs			
Taxes		--	.14
Use of machinery		--	90
Overhead		--	3.46
	Total	--	4.50
Total cost exclusive of interest		--	27.60
Interest at 6%		--	8.14
Total cost including interest		--	35.74
Quantity of nuts at 30¢ a lb. required to cover cost including harvest			
Exclusive of interest	pounds	103	--
Inclusive of interest	"	132	--

Table 7. -- Cost of preparing an acre of land for corn, 1860-1930.

Year	Spreading fertilizer	Plowing	Disking	Harrowing	Planting	Cutting and shocking	Husking
1860	2.40	2.75	--	1.25	.75	1.30	2.00
1880	2.80	3.00	--	1.25	.85	1.40	2.00
1900	3.00	3.20	.75	.60	1.00	1.75	2.50
1910	3.00	3.20	.80	.65	.50	2.00	3.00
1930	3.50	2.50	.90	.40	.75	2.50	3.25
Total	14.70	14.45	2.45	4.15	3.85	8.95	12.75

Table 8 -- Percentage of citrus fruit moved each month by the California Fruit Growers Exchange, November, 1910, to October, 1925

Month	Average percentage shipped			Month	Average percentage shipped		
	1910-1915	1915-1920	1920-1925		1910-1915	1915-1920	1920-1925
November	5.5	6.1	5.6	May	14.2	12.9	12.2
December	9.1	7.8	7.4	June	10.0	10.9	11.4
January	7.9	6.3	7.5	July	6.1	7.9	8.3
February	8.6	7.9	7.3	August	3.7	5.8	6.5
March	13.4	11.8	11.1	September	3.3	4.5	6.0
April	14.8	12.7	11.2	October	3.4	5.4	5.5

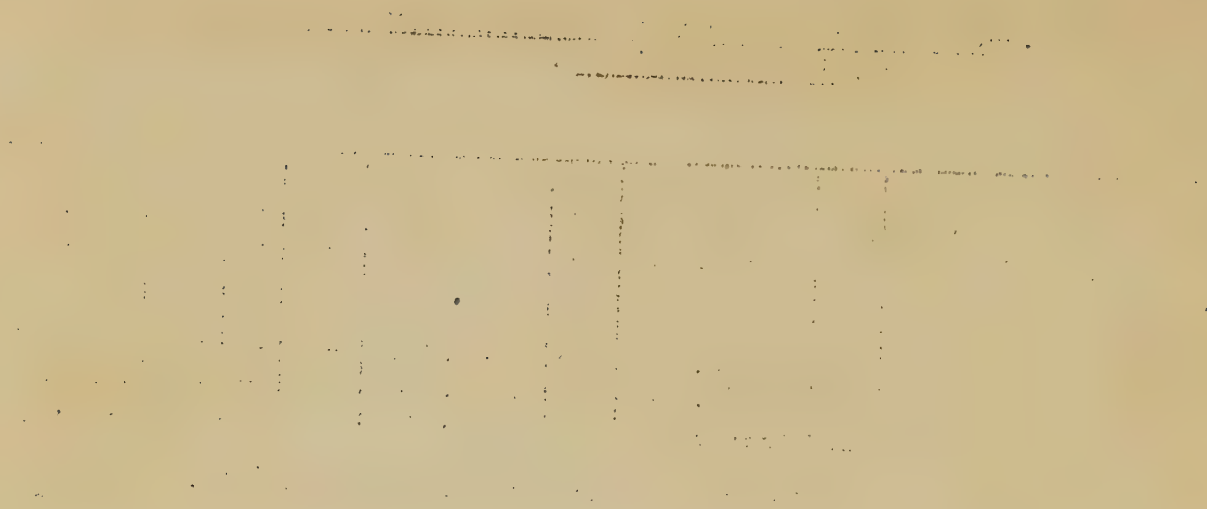
Table 9.--Statistical summary of results of cooperative extension work, 192-

Item	Reported by county agents		Reported by home demonstration agents		Reported by club agents[1]		Total	
	Agents reporting	Project or line of work	Agents reporting	Project or line of work	Agents reporting	Project or line of work	Agents reporting	Project or line of work
	Number	Number	Number	Number	Number	Number	Number	Number
Communities in counties	2,227	41,939	919	24,380	129	3,758	3,275	70,077
Communities with extension program	1,873	23,213	772	13,377	117	2,708	2,762	39,298
Farm visits made	2,238	1,200181	180	38,481	119	42,374	2,537	1,281,036
Different farms visited	2,090	569,698	176	18,053	108	22,136	2,374	609,887
Home visits made	953	163,084	906	234372	95	22,575	1,954	420,031
Different homes visited	928	89,683	880	120,623	86	11,826	1,894	222,132
Office calls	2,216	3,374,761	907	556,137	121	55,010	3,244	3,979,908
Percentage of time in field	2,234	66	919	69	131	64	3,284	67
Percentage of time in office	2,237	34	919	31	131	36	3,287	33
Individual letters written	2,232	2,677,577	917	655,185	131	112,118	3,280	3,444,880
Demonstration meetings	2,019	110,191	820	123,614	120	11,422	2,959	245,227
Attendance at demonstration meetings	2,023	2,720,660	799	2376,721	120	201,063	2,942	5,298,444
Number of all meetings	2,019	235,020	820	161,348	120	24,369	2,959	420,737
Attendance	2,023	10,868,483	799	4,399,750	120	749,538	2,942	16,017,771
Soils								
Adult demonstrations	1,466	34,502	1	10	4	38	1,471	34,550
Farms following advice in use of commercial fertilizer	1,518	169,618	2	125	4	316	1,524	170,059
Farms using lime and limestone on advice	1,280	63,702	--	--	1	17	1,281	63,719
Farms taking better care of farm manure	1,092	60,225	--	--	2	518	1,094	60,743
Farms plowing under green-manure crops	1,368	57,165	1	10	3	254	1,372	57,429
Different farms adopting better practice	1,971	322,065	2	125	7	819	1,980	323,009
Sweetclover:								
Adult demonstrations	684	7,059	--	--	2	31	686	7,090
Junior demonstrations	4	27	--	--	-	--	4	27
Farms planting selected seed	463	12,300	--	--	-	--	463	12,300
Farms inoculating for this crop	543	11,147	--	--	1	2	544	11,149
Different farms adopting better practices	1,078	30,014	--	--	2	109	1,080	30,123

Table 9.--Statistical summary of results of cooperative extension work, 192 --Continued.

Rodents:								
Adult result demonstrations	150	1,545	7	156	--	--	157	1,701
Pounds of poison bait used	141	129,929	3	24	--	--	144	129,953
Marketing:								
Membership in cooperative marketing association	469	818,254	86	38,894	--	--	555	857,148
Value of products marketed	489	295,986,825	91	1,707,515	--	--	580	297,694,340
Value of supplies purchased	320	42,924,560	18	45,012	--	--	338	42,969,572

[1]Includes a small amount of work in counties without extension agents, reported by State club leaders.



Construct Table 10 from the following:

A comparison is made of the time and labor required at 123 milk plants, 91 of which received the milk in cans from trucks, which is the method used at the plants previously described (method 1); 24 of which received the milk from tank trucks or tank cars (method 2); and 8 of which received it in cans from cars switched alongside the receiving platforms(method 3).

An average of 1,540 gallons of milk per hour and 197 gallons per man-hour was received at the plants using method 3, 1,565 gallons of milk per hour and 384 gallons per man-hour at the plants using method 1, and 1,834 gallons of milk per hour and 1,309 gallons per man-hour at the plants using method 2. The plants that received this milk from tank trucks or tank cars handled more than three times as much per man-hour as the average amount handled at the plants using method 1 and more than six times the average amount handled by the plants using method 3.

The 91 plants, which employed an average of 3.9 men per plant, received an average of 7,328 gallons of milk daily in 4.6 hours. The 24 plants, which employed an average of 1.5 men per plant, received an average of 8,102 gallons of milk daily in 4.4 hours. The 8 plants, which employed an average of 8.4 men per plant, received an average of 9,145 gallons of milk in 5.9 hours. It required 19.1 man-hours to receive the milk by the plants using method 1, 6.2 man-hours by the plants using method 2, and 46.5 man-hours by the plants using method 3. One hundred gallons of milk were received in 3.3 minutes by the plants using

method 2, in 3.8 minutes by the plants using method 1, and in 3.9 minutes by the plants using method 3. The method of receiving from tank trucks or cars requires much less labor than does receiving the milk in cans, as the washing and steaming of the tanks are small tasks as compared with the washing of the large number of cans. Furthermore, the milk is transferred from the tanks to the weigh can or receiving tank by pump, gravity, or air pressure, so that no handling is required.

Elementary Statistics

Rules and Problems in Percentages

The Terms used in percentage are: Base, Rate, Percentage, Amount, Difference.

The Base is the number upon which the percentage is calculated.
The Rate is the number of hundredths to be taken, and hence is always an abstract number.
The Percentage is the result obtained from taking the required per cent of the base, and is always of the same denomination as the base.
The Amount is the sum of the base and the percentage, and is always of the same denomination as the base.
The Difference is the base minus the percentage, and is always of the same denomination as the base.
Two of these terms must be given to find a third.

Rule 1. - To find the percentage when the base and the rate are given, multiply the base by the rate expressed as a common fraction, or a decimal.

Rule 2. - To find the rate when the base and the percentage are given, divide the percentage by the base.

Rule 3. - To find the base when the percentage and the rate are given, divide the percentage by the rate.

Rule 4. - To find the base, divide the amount by 1 plus the rate, or the difference by 1 minus the rate.

A horse cost $250 and was sold at a loss of 18%. What was the selling price?

A farmer had 1000 bushels of wheat; he sold to one man 33-1/3% of it, and to another 25% of the remainder. How many bushels did he sell, and how many bushels remained unsold?

A farmer had a flock of 800 sheep. He sold 1/5 of them to one man, 200 sheep to another, and 25% of the remainder to another. What % of his flock had he remaining?

A house cost $5000 and rents for $600 per year. The taxes and repairs are $150 per year. What per cent does it pay on the investment?

A owned 66-2/3% of a ship and sold 25% of his share for $2000. What was the value of the ship?

B lost $6, which was 15¾% of what he had left. How much had he at first?

A banker, after paying 60% of his debts, found that $3000 would pay the remainder. How much did he owe at first?

(over)

An army lost 11% of its men in battle, and had left 4450 men. How large was the army at first, and how many men were lost in battle?

There are enrolled in school 125 pupils, of which 55 are boys and the remainder girls. What per cent of each?

A flock of sheep increased 300, or 15%, in one year. How many sheep in the flock at first?

A gained 5% in weight in January, and lost 5% in weight in February. What is the relation of his weight March 1 to his weight January 1?

A man owning 40% of a ship sold 33-1/3% of his share for $2000. What is the value of the ship?

A mill is worth 4% less than a farm, and the farm is worth 16% more than a house and lot. The owner of the house and lot traded it for a ½ interest in the mill, and thereby lost $412. Find the value of each.

The amount paid the teachers in a district is $3740; what amount of tax must be assessed for the teachers' fund if the cost of collecting the tax is 2%?

A book bought at $1.25 was sold at a profit of 30%. For how much was it sold?

The total value of Argentine exports declined from 454 million gold pesos in 1929 to 341 million gold pesos in 1930, a decline of _____ per cent.

I invested 45% of 33-1/3% of my money, and had $1500 left. How much had I at first, and how much did I invest?

Statistical Tables

A statistical Table is an arrangement of data in rows and columns according to some definite classification. Classification is the combination of similar or like datum into groups. The purpose of a Table is to arrange data in a condensed form to facilitate comparisons. A Table consists of (1) a number, (2) title, (3) source, (4) legend, (5) rulings, (6) stubs, (7) captions, and (8) contents.

(1) The purpose of the number is to make reference to the Table simple and clear. In a large Table columns and rows should be numbered or lettered also.

(2) The title should be clear, concise and catchy. It should tell exactly what is in the Table, when the data were gathered, or events occurred, where data were collected from, and if possible, the classifications made. If title is too long use a main title and a sub-title with different sized lettering to emphasize important parts.

(3) The source of the data should be given. If you are the original collector of the data; it should be noted in the article in which the Table appears rather than in a footnote. Otherwise, a footnote giving the source should accompany each Table.

(4) The legend should tell in what units the data in the Table are expressed. If all of the data in the Table are expressed in the same unit, then the legend may be placed above the top ruling or directly under the title. If different units are used for different stubs or captions, the legend should appear at the head of the column or at the left of the row.

(5) Each columnar line and each space between rows in the Table should have a definite purpose. There are several styles of ruling Tables, but one should decide upon the style best suited for their purpose and then follow it rigidly.

(6) and (7) Stubs and captions should be arranged in some definite order, i.e. chronological, alphebetical, geographical, or according to some natural division of the subject matter. The stubs and captions should adhere to the definitions or limitations expressed by the stubs and captions respectively, and if repeated should always be arranged in the same order each time. They should be brief, but long enough to be self-explanatory. They must be rigidly defined.

(8) Only data that are closely related should be shown in the same Table. All possible repetition of figures in a Table should be avoided. Every Table should be as simple as possible. Simplification leads to a better understanding of the contents and brings out more clearly the comparisons that are intended. The data should be arranged so that comparisons to be most frequently made can be most easily made.

Elementary
Statistics

TABLE CONSTRUCTION

1. Criticise and rewrite the following titles:
   (a) "Amounts paid to collectors from the proceeds of legal actions, etc., during the fiscal year."
   (b) "Aggregate collections made and reported to the Commissioner of Internal Revenue by collectors of the several districts and by stamp agents in the Philippine Islands during the calender year ended with December 31, 1930."

2. (a) How big should you make a table?
   (b) How would you construct a table which is too large to go on one page? Illustrate.
   (c) Where would you place averages, ratios or percentages in a table? Illustrate.

3. Plan the stub and caption for a table showing:
   (a) The number of insane in hospitals in the United States, by States, January 1, 1920, 1925 and 1930.
   (b) The percentage of population under 15 years of age in the United States, by sex and color, January, 1920 and 1930.

4. A manufacturing concern employs five salesmen, each with an exclusive territory in which to sell the products of the company. The executive of the company wishes to compare the work of the various salesmen at the close of each month and at the end of the year. Draw up a table to give this information assuming that each of the five salesmen handle only three items, "A", "B" and "C".

5. The city officials of Baltimore ask you to present in table form data based upon the returns of several thousand questionnaires. The officials would like the tabulation in such a form that comparisons may be made between various sections of the city, their respective population, and available transportation facilities.

6. Criticise:

Causes of Accidents Resulting in Infection

Causes of Accidents	Total	Fatal	Amputations	Infected Cuts	Infected Bruises	Infected Burns	Infected Eyes
Nails in Floor	721	5	4	511	102	53	45
Total	9,299	25	10	8,738	346	100	80

8. Criticise:

Frequency table showing
Classified Weekly Wages for Employees in all Manufacturing Industries, Massachusetts, 1930

Wage Groups	Number and Per Cent of Employees Receiving Specified Amounts	
	Number	Per Cent
Under $3		
$3 but under $4		
$4 but under $6		
$6 but under $10		
$60 and over		
Total		

9. Prepare a table to show the farms of Brazos County, Texas, according to size, type of tenure, color of operator and amount of net income per farm using the two classes of income, "less than $800 net income" and "over $800 net income". Place emphasis upon these four items in the order named. Give the table an appropriate title. The data to be placed in the table are as follows:

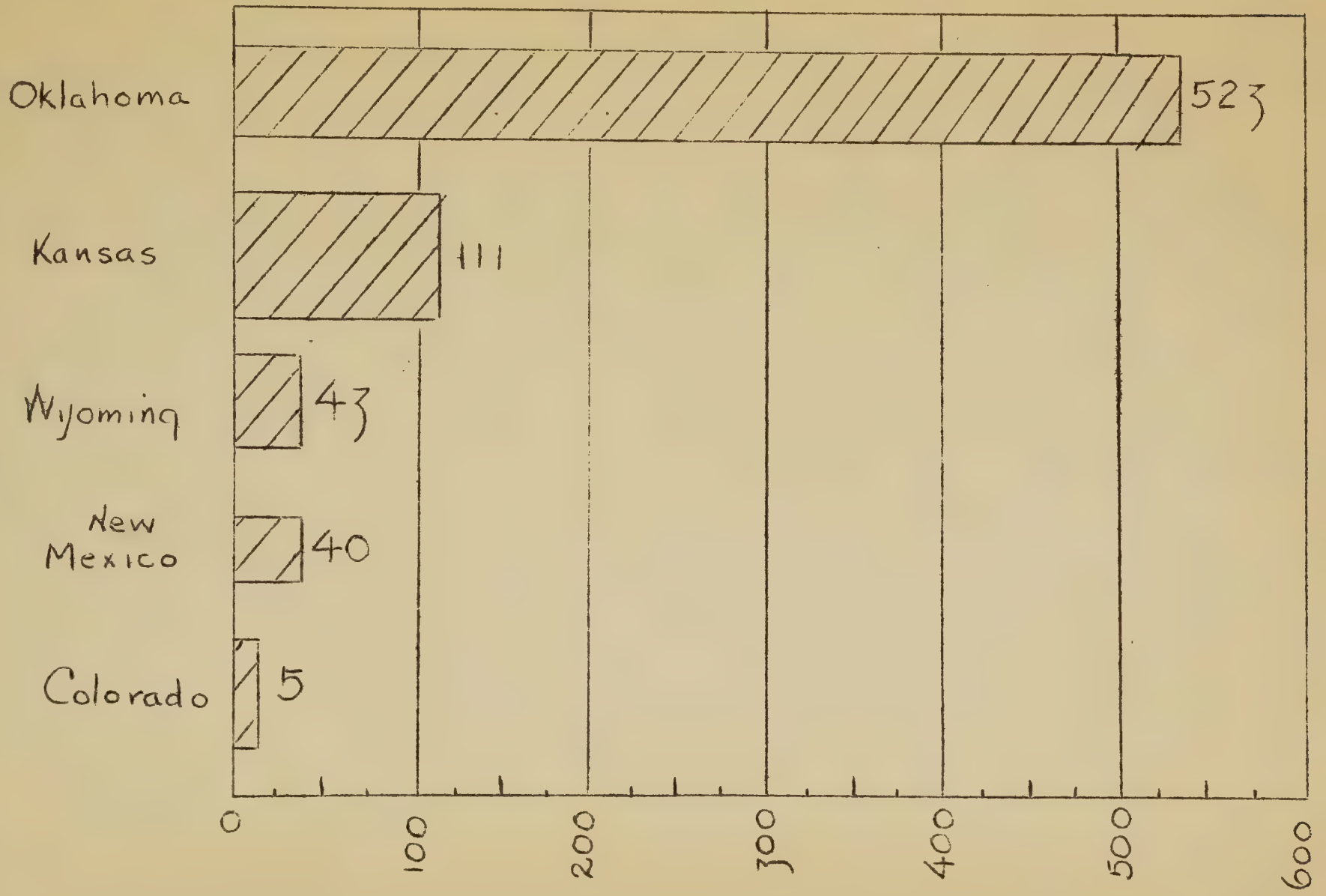
THE NUMBER OF WHITE FARMERS IN BRAZOS COUNTY, TEXAS
Size of Farm (Acres)

	Under 40	40-79	80-159	160-319	320-640	Over 640
TENANTS						
Less than $800	98	280	106	11	6	3
Over $800 Income	14	120	62	130	48	6
OWNERS						
Less than $800	62	200	122	9	3	0
Over $800 Income	22	114	224	92	36	14

THE NUMBER OF COLORED FARMERS IN BRAZOS COUNTY, TEXAS
Size of Farm (Acres)

	Under 40	40-79	80-159	160-319	320-640	Over640
TENANTS						
Less than $800	156	382	54	5	2	0
Over $800 Income	4	64	16	20	1	0
OWNERS						
Less than $800	124	224	70	3	1	0
Over $800 Income	14	90	76	16	2	0

DAILY AVERAGE PRODUCTION OF CRUDE OIL IN THOUSANDS OF BARRELS IN THE FIVE OIL PRODUCING STATES OF THE TENTH FEDERAL RESERVE DISTRICT AS ESTIMATED BY THE AMERICAN PETROLEUM INSTITUTE IN MONTH OF MARCH, 1931



FJH-10.10.34

QUESTIONS ON THE COLLECTION OF STATISTICAL DATA

1. Outline briefly the procedure necessary to carry on a statistical investigation.

2. Enumerate the factors or elements which determine whether a census is a good one or not.

3. (a) An agricultural experiment station is interested in the marketing of tomatoes. The tomato growers contend that the use of fertilizer rich in potash tends to toughen and strengthen the skin of the tomato so that it may be shipped long distances. You have been chosen to plan an investigation which will verify or contest the opinion of the growers. Explain BRIEFLY the steps necessary in the preparation for the actual work of investigation.

   (b) Indicate briefly what method of procedure would be most appropriate and practical for an investigation into:

      1. Buying habits of women customers.

      2. Causes of juvenile delinquency.

      3. Changes in the cost of living in Washington, D. C. from 1933 to 1934.

      4. A statistical measurement of recreational facilities in a rural, sub-urban, and urban community.

      5. Relation between rent received and land values.

   (c) Discuss briefly the problems you would expect to encounter in constructing a series of monthly average prices for a particular commodity in a particular market from actual daily sales in that market.

   (d) To what sort of problems are statistics and statistical methods inappropriate?

4. (a) What tests whould a statistician apply to secondary data before using them?

   (b) Name at least ten general sources of secondary statistical data.

5. Name and explain the advantages and disadvantages of the various methods of collecting data from original sources.

6. Explain the outstanding characteristics of a well-prepared questionnaire.

7. Criticise the following questions found on questionnaires:

   (a) "Did you read this magazine from cover to cover?"

   (b) "What suggestions do you make for development of college spirit at your institution?"

   (c) "How much cotton, corn, wheat, oats, etc., have you sold in the past year?"

   (d) "How large is your farm?"

   (e) "How many members of your family have attended college?"

   (f) "What is the trend of the death rate in your community?"

   (g) "Name the biggest man in your city".

   (h) "What is the rate of labor turnover in your factory?"

8. Formulate a good usable definition of the following statistical units:

   (a) A "residence", for use in studying housing conditions in Washington, D. C.

   (b) An "Improved farm", to be used in a farm survey in Maryland.

   (c) A "family", to be used in a cost-of-living analysis.

   (d) A "retail merchant", for use in a distribution study.

   (e) An "insane person".

   (f) An "establishment", for use in a census of manufactures.

   (g) A "criminal".

9. (a) Distinguish briefly between the library, field and correspondence methods of obtaining data.

   (b) "Data which are primary in the hands of one statistician may be secondary in the hands of another". Explain and give examples.

10. "Editing of schedules has in mind four purposes". Explain.

11. Name the requisites of a good tabulation sheet or schedule.

Required Materials for "Elementary Statistics"

1. Text:

   (a) Mills, F. C., "Statistical Methods Applied to Economics and Business". This book may be secured from several sources, including Washington book stores (second-hand or new book stores), from friends, or by ordering a new copy through the instructor, who in turn will order it through the Bureau of Agricultural Economics Library.

2. Suggested Supplies: (may be secured at a book or stationery store)

   (a) A heavy covered notebook in which to file mimeographed material and take notes on lectures and on material contained in the textbook or other books.

   (b) Pad of graph paper (10 spaces to an inch, preferably blue in color).

   (c) Five sheets of 2 cycle semi-log paper.

   (d) Fountain pen, pencil, ruler, ink and pencil eraser, and art gum.

Class Regulations

3. (a) The class meets twice each week from 4:50 to 5:50 P.M. Roll will be called, and attendance, non-attendance or tardiness will be recorded.

   (b) If a student is unable to attend a class, the assignment may be obtained by phoning Mr. B. R. Stauber, Extension 2065, or Mr. F. J. Hosking, Extension 2241. Both phones are on the District 6350.

   (c) All assigned problems must be completed before a grade will be issued for the course. Each individual piece of work will receive a full grade when it is completed and submitted by the

designated time, but a deduction of one letter will be made for each week an assigned problem is late. A reasonable excuse will be accepted if work cannot be handed in on time.

(d) The following letters will be used to indicate the quality of work submitted:

A - Excellent
B - Good
C - Fair
D - Pass
E - Condition
F - Failure

(e) 5, 10, or 15 minute examinations will be given from time to time without previous announcement. An examination of an hour or more in length will be given at mid-semester and/or at the close of the semester.

(f) Grades on the assigned problems make up about one half of the final grade, recitations and short-time quizzes about one fourth, and the mid-semester and final examinations about one fourth.

2. VARIABLES AND STATISTICAL UNITS

(Taken largely from Day, "Statistical Analysis", Chapter 2)

I. The general purpose of statistical inquiry is the analysis and interpretation of a variable or variables. (Day, page 10.) (Refer to definitions of statistical method in previous outline.)

II. Statistics of attributes and statistics of variables. (Yule, first several pages of chapter entitled, "Notation and Terminology".)

III. Variables.

a. Definition.

1. A variable is anything that exhibits discernible differences of magnitude or of number, i.e., one's pulse, weight, stature price of eggs, etc.

A. Single object at different times.

B. Single objects at different places.

C. Repeated measurements of same object.

D. Single observed element in a number of related objects.

E. Differing frequencies of occurrence.
(Day, pages 10-24; see also Jerome, pages 36-37; Mills, page 14; Chaddock, page 54.)

b. Determination of variables.

1. By tabulation of individual cases.

2. By specific measurement.

3. By computation. (Sometimes called derived variables.)

A. Simple averages.

B. Compound variables. (Sometimes called relatives.)

C. Rates, ratios, and coefficients. (See Jerome, chapter 8; also Crum and Paton, "Economic Statistics", pages 152-154.)

c. Types of variables.

1. Continuous, approximately continuous, and discrete. (Chaddock pages 72-73.)

2. Independent and dependent.

3. Related and unrelated.

IV. Statistical units.

a. Definition.

b. Necessity of a clear-cut definition of the statistical unit.

c. Types of units; (Day, pages 18-23.)

1. Natural objects.

2. Produced commodities.

A. Simple.

I. Physical.

II. Pecuniary.

B. Compound.

C. Abstract.

3. Institutional objects.

d. No serious problem arises in defining natural objects, but considerable difficulty is encountered in produced and institutional objects.

THE COLLECTION OF STATISTICAL DATA

I. Predetermined plan of procedure. (Jerome, pages 290-292; Secrist, revised edition, pages 47-53; Chaddock, pages 371-374).

a. Purpose of the investigation.

b. Determination whether the problem as formulated lends itself to statistical treatment.

c. Examination of available literature and previously collected data. (Secondary data).

1. Types of data necessary for the problem, analysis or solution.

2. Are the data available in suitable form?

3. Are the data adequate for the problem?

4. Have the data the required degree of accuracy, consistency, and comparability?

5. Proportion of problem which can be solved with secondary data; with primary data. Consideration must be given to time and money.

d. Choice of methods to be used in obtaining facts from secondary sources; from primary sources.

e. Selection of working force to collect secondary data; to collect primary data.

f. Preparation of the necessary forms and instructions for their use in the field by the working force.

g. Instructions for editing the schedules after they are returned by the enumerator or other informant to the statistician.

II. Methods of collecting data. (Note references below.)

a. Library method.

b. Field method.

c. Office or correspondence method.

1. Compulsory.

2. Cooperative.

III. Library method.

a. Definition: The library method involves the use of secondary sources, of compilations of facts prepared by some previous investigator. (Jerome, pages 292-294; Chaddock, pages 392-394; Crum and Patton, Chapter III; Day, Appendix B, pages 389-392).

b. Tests to be applied to secondary data before they are used. (These tests listed or discussed in most texts.)

1. Are the data homogeneous throughout the period for which they are available or the period for which they are to be utilized?

2. Is the validity of the data affected by an intentional or an unintentional bias?

3. Are there other elements of inaccuracy, due, for example, to the imperfection of the sources or of the methods used in collecting the data?

4. To summarize, are the data as adequate for the problem in hand as data which might be obtained from primary sources?

IV. Field method. (Jerome, pages 294-299; Chapin, "Field Work and Social Research", Chap. V.)

a. Source of data.

1. Official records: birth, death, and marriage certificate records.

2. Private records: private accounts, cost-of-living budgets, etc.

3. Individual unrecorded knowledge. (U.S.D.A. Bulletin 1020).

4. Direct observation.

b. Personnel of the investigating or working force.

1. Personal character of workers.

2. Standards for efficiency.

A. Study of conditions under which inquiry is to be made.

B. Definition of the statistical units.

I. Appropriateness for purpose of study.

II. Clarity.

III. Measurability.

IV. Comparability.

3. Forms to be used.

c. Scope of inquiry.

1. Complete.

2. Partial.

V. Office or correspondence method. (Jerome, pages 299-305; Chaddock, pages 374-395; Day, pages 383-388; Crum and Patton, Chapter IV; Lovitt and Holtzclaw, Chap. II).

a. Definition: The office method involves the substitution of inquiry by mail or by other means for field investigation.

b. Types of questionnaire studies.

1. Questionnaire - schedules filled in by a reporter.

2. Questionnaire - schedules filled in by an informant.

3. Questionnaire - schedules mailed to and filled in by an informant and later checked by an enumerator.

c. Questionnaire schedule filled in by a reporter.

1. Success of this method depends upon:

A. Personal ability of the individual making the interview.

B. Personal qualities of interviewer -- tact, diplomacy, courage, intellectual curiosity, integrity, etc.

C. Degree to which he understands:

I. The problem upon which he desires information.

II. The reactions of those he interviews.

D. Accuracy with which he:

I. Interprets information supplied.

II. Records and remembers facts submitted.

E. Form of record upon which answers are put.

d. Questionnaire schedule filled in by an informant.

1. Success of this method depends upon:

A. Locating and addressing questionnaire to those who have access to the facts.

B. Definition and precise statement of what is wanted.

C. Data requested should be as far as possible a matter of record rather than of opinion.

D. Formulating the questionnaire in such a manner that the units in which the data are measured are:

I. Same as those in current use.

II. Simple rather than being composite or expressed as ratios.

E. Ability of questionnaire to overcome indifference and reluctance of the informant to give information which may be:

I. Confidential.

II. Difficult or costly to assemble.

III. Of value to competitors.

c. Form and nature of questionnaire.

1. Explanatory letter should accompany the questionnaire.

A. The letter should state the purpose for which information is being collected.

B. It should quote the authority or auspices under which investigation is being made. Assurance should be given that inquiries are made according to the provisions of the law, or if voluntarily taken, with hope of throwing light on some problem.

C. It should give some inducement for an immediate reply.

D. It should promise a copy of results of the investigation to the informant.

E. Promise should be made to the informant that no individual data will be disclosed -- only totals and averages will be published.

2. Questionnaire should be accompanied by an addressed and stamped (or franked) envelope.

3. Form of questionnaire.

A. If possible the questionnaire should be of a convenient size to carry and send without folding.

B. If possible it should conform in size to some one of the standard filing devices, i.e., 5 x 8 inches, 8½ x 11 inches.

C. Quality of paper should be such as to permit use of ink, sufficiently durable to stand handling and sorting without replacement.

D. Careful attention should be given to the spacing and kind of type used in printing.

E. Number or letter each question or item in schedule to render easy identification and facilitate tabulation.

F. Experiment with sample questionnaires in a restricted locality to determine coverage, omissions, errors, etc.

G. If possible, it is desirable to furnish a duplicate for the records of the informant.

4. Form and nature of questions.

A. Units of measurement should be clearly indicated, accurately defined, and should conform with common usage. Units may appear at beginning, in body, or at end of questionnaire. For interview questionnaire, at beginning; for mail questionnaire, in proximity to entries to which they relate. Explanation of units should include a statement of the degree of numerical accuracy expected in answers.

B. The general appearance of questionnaire should not be crowded. Allow sufficient space for completion. Related questions should be grouped together.

C. False or inaccurate answers should be guarded against by having questions, so far as possible, corroboratory.

D. Questions should be simple, definite, clear, courteous, diplomatic, easy to answer, not inquisitorial. Inquirer should prevent evasive or double interpretation type of questions. Use "yes" or "no", "number" or "amount", or "check" questions. Use few questions; the fewer the questions, the greater the percentage return.

E. As a rule, the making of arithmetical calculations such as totals, ratios, percentages, etc., should be reserved for the statistical organization.

F. Rulings and columner arrangement should be simple and definite to prevent misplacing of items.

G. In every possible way reduce the task of the informant to a minimum in order to receive maximum returns. Ask informant to sign the questionnaire. This tends to increase the accuracy of his report.

f. Distribution of mail questionnaires.

1. To whom shall questionnaires be sent.

A. Method of sampling to be used.

2. All questionnaires to be sent out at the same time.

3. Determine closing date after which questionnaires will not be used. This aids in tabulating.

g. Return of mail questionnaires.

1. Success will probably attend project if

A. Informant is impressed with necessity of his completing and returning the questionnaire.

B. Mandatory power possessed.

C. Degree of cooperation between informant and inquirer is fairly high.

h. Editing of mail questionnaires.

1. Necessity of checking, editing and revision.

A. Editing is reduced to a minimum if questionnaire is properly made out.

2. Purpose of editing is to check:

A. Accuracy.

B. Consistency.

C. Uniformity of expression.

D. Completeness.

3. Editing.

A. As a rule the editor accepts as final the replies entered upon the schedule, but if there is a presumption of error he attempts to verify the answer.

B. Errors should be noted in a distinctive manner or color on the questionnaire.

C. Errors of omission, additions, false entry, and confusion of items may be readily corrected by editor. Alterations should be made only in case of an unmistakable error.

D. The degree to which omissions may be tolerated should be determined before beginning of editing. Study remainder of schedule and determine whether it is a representative sample.

i. Editing sheet.

1. Function.

A. To classify, arrange, and summarize in easily accessible form the answers to the questions with which the inquiry is concerned.

2. Method.

A. Preliminary: Analyze the factors in the problem investigated. Intelligent tabulating begins with making editing sheet such as to economically utilize results of inquiry.

B. Process of tabulation.

I. Form of editing sheet. Usually columnar in type divided into spacing with proper headings, rows numbered identically with questionnaire number, space for computing percentages, ratios, etc.

II. Tabulation in table.

III. Mechanical tabulations.

a. Various devices.

(1) Punch card.

(2) Key punch.

(3) Sorting machine.

(4) Total and listing machine.

j. Necessary budget for financing of investigation.

References

Bailey, W. B., and Cummings, John, "Statistics", Chap. III, pages 8-16, Chap. IV, pages 17-25.

Chapin, F. S., "Field Work and Social Research", Chap. VII, pages 148-191.

Rugg, H. O., "Statistical Methods Applied to Education", Chap. II, pages 39-56.

Secrist, H., "An Introduction to Statistical Methods", Revised Edition, Chap. II, III, IV, and V, pages 22-123.

Jerome, H., "Statistical Method", Chap. XVI and XVII, pages 290-326; also Chap. 2 (The sampling process).

Day, E. E., "Statistical Analysis", Appendix A and B, pages 383-392.

Chaddock, R. E., "Principles and Methods of Statistics", Chap. XIV, pages 371-396.

Pearl, Raymond, "Medical Biometry and Statistics", Chap. III.

Riegel, R., "Elements of Business Statistics", Chap. VI and VII.

Bowley, A. L., "An Elementary Manual of Statistics", part I, Chap. 7 and 8.

Crum and Patton, "Economic Statistics", Chap. III and IV, pages 22-38.

Lovitt and Holtzclaw, "Statistics", Chap. II and III.

1. NATURE AND USES OF STATISTICAL METHOD

I. Meaning of the word, "statistics".

a. Origin of the word, "statistics".

1. "Statistics" appears to be derived from the Italian, "ragione di stato", practical politics, and "statista", statesman.

b. Historical development of the word, "statistics".

1. "Statistics is the science that teaches us what is the political arrangement of all the modern states of the known world". This is considered one of the earliest known definitions of "statistics". (Noted in "The Elements of Universal Erudition", J. F. Bielfeld, 1770. From Yule, "An Introduction to the Theory of Statistics", page 1.)

2. A wider definition appears in the preface to "A Political Survey of the Present State of Europe", E. A. W. Zimmerman, 1787. He states that "that branch of political knowledge which has for its object the actual and relative power of the several modern states, the power arising from their natural advantages, the industry and civilization of their inhabitants, and the wisdom of their governments, has been formed.....into a separate science". (From Yule, page 1.)

3. The early exposition of the noteworthy characteristics of the state, or "statistics", was preponderantly verbal. In time "statistics" acquired a more narrow meaning, viz., the exposition of the characteristics of a state by numerical methods.

4. The word "statistics" was transferred to those data with which it operated, such as vital statistics, economic statistics, etc.

5. Today, the term statistics is used somewhat indiscriminately to refer to any one of three concepts - statistics, statistical method, and statistical theory.

II. Explanation of the modern concepts of "statistics".

a. Statistics or statistical facts.

1. Quantitative data affected to a marked degree by a multiplicity of causes. (Yule, page 5.)

2. Aggregates of facts numerically expressed. (Secrist, "An Introduction to Statistical Method", page 10.)

b. Statistical method or technique.

1. The science of counting. (Bowley, "Elements of Statistics", 4th edition, page 3.)

2. The elucidation of quantitative data affected by a multiplicity of causes. (Yule, page 5.)

3. The term "statistics", when used to designate a branch of study implies an exposition of certain methods employed in presenting and interpreting the numerical aspects of a given subject. (Davies, "Introduction to Economic Statistics", page 3.)

4. The method of judging collective natural or social phenomena from the results obtained by the analysis of an enumeration or collection of estimates. (King, "Elements of Statistical Method", page 23.)

5. The principles of the collection and the analysis of numerical data and the accurate and effective presentation of such data and the results of their analysis. (Jerome, "Statistical Method", page 2.)

c. Statistical theory.

1. The exposition of statistical methods. (Yule, page 5.)

2. Mathematical and economic principles which underlie and justify statistical methods. (Jerome, page 1.)

III. The uses of statistical method.

a. Pre-eminent causes for modern importance of statistics.

1. As listed by Jerome, pages 3-5.

A. This is an age of large numbers.

B. Modern science demands its theories be based on facts.

C. Business is becoming more and more a matter of scientific procedure.

2. As listed by Day, "Statistical Analysis", pages 6-8.

A. Growing need of more precise measurement in human affairs.

B. Spread of the modern scientific spirit.

C. Statistical method is the only science able to study social complexes.

b. Application of statistical method.

1. As listed by Jerome, pages 5-7.

A. Analysis of social phenomena.

B. Economic conflicts.

C. Commercial and political use.

I. Public administration.

II. Internal and external business statistics.

2. As listed by Chaddock, "Principles and Methods of Statistics", pages 31-39.

A. Statistics in the service of social sciences.

I. Dependence of the social sciences on statistics.

B. Service of statistics to economics and sociology.

C. Statistics in the service of education.

3. As listed by Mills, "Statistical Methods", chapter I.

c. Requisites for success in statistical study and practice.

1. Constructive imagination, intellectual integrity, capacity for careful work, and knowledge of subject matter. (Jerome, pages 9-11.)

2. Good judgment, broad knowledge and experience, and common sense are the most valued possessions of a statistician. (Chaddock, page 31.)

References

(These readings are suggested for the purpose of presenting different points of view from which writers have approached the study of statistical methods.)

Mills, F. C., "Statistical Methods", chapter 1.
King, W. I., "Elements of Statistical Method", part I.
Chaddock, R. E., "Principles and Methods of Statistics", part I, pages 3-39.
Pearson, Karl, "The Grammar of Science, Physical", part I, 3rd edition, chap. 1.
Yule, G. U., "An Introduction to the Theory of Statistics", Introduction.
Jerome, Harry, "Statistical Method", chapter 1.
Mayo-Smith, R., "Statistics and Sociology", chapters 1, 2, and 3.
Whipple, G. C., "Vital Statistics", 2nd edition, chapter 1.
Pearl, Raymond, "Medical Biometry and Statistics", chapter 1.
Jones, D. C., "A First Course in Statistics", chapter 1, page 4.
Burgess, R. W., "Introduction to the Mathematics of Statistics", chapter 1.
Lovitt, W. V., and Holzclaw, H. F., "Statistics", chapter 1.

QUESTIONS ON
THE NATURE AND USES OF STATISTICAL METHOD

1. Outline the method of forming a judgment of ideas and concepts. What proportion of this process is quantitative?

2. (a) Compare the development of the methods of statistics with the growth of nations.

   (b) How can statistics be used as an aid in the development of a powerful nation?

3. Criticize:

   (a) "Statistics may, for instance, be called the science of counting".

   (b) "The term 'statistics' when used to designate a branch of study implies an exposition of certain methods employed in presenting and interpreting the numerical aspects of a given subject".

4. Distinguish between statistical facts, statistical theory and statistical methods. With which are we dealing in this course?

5. Distinguish between accounting and statistics.

6. Name and explain the reasons for the modern importance of statistics to students of agriculture and economics.

7. What evidence is there to support the contention that there is an increasing need for the development of statistical science in agriculture?

8. What concrete illustrations can be given of the use of statistical method in agriculture? In business? In sociological study? In the administration of the laws of economic significance? In labor conflicts?

9. What relation is there between the statement that intellectual integrity is an essential for statistical practice and the frequently heard assertion that "anything can be proved with figures"?

10. What personal qualities are requisite for success as an agricultural statistician? Are the same qualities necessary in a student of statistics? Why?

TEXT AND REFERENCES

Text: Mills, F. C., "Statistical Methods Applied to Economics and Business".

Suggest References:

American Telephone and Telegraph Co., Statistical Methods Series No. 1, issued by Chief Statistician; "Introduction to Frequency Curves and Averages"; "Statistical Analysis and Projection of Time Series"; also "Graphical Method of Smoothing a Series of Frequency Curves".

Bailey, W. B., and Cummings, John, "Statistics".

Bowley, A. L., "Elements of Statistics"; also, "An Elementary Manual of Statistics" (Various editions).

Brinton, W. C., "Graphic Methods for Presenting Facts".

Brunt, David, "The Combination of Observations".

Burgess, R. W., "Introduction to the Mathematics of Statistics".

Chaddock, R. E., "Principles and Methods of Statistics".

Chambers, G. G., "An Introduction to Statistical Analysis".

Chapin, F. S., "Field Work and Social Research".

Coolidge, J. L., "An Introduction to Mathematical Probability".

Crum, W. L. and Patton, A. C., "An Introduction to the Methods of Economic Statistics".

Davenport, C. B., "Statistical Methods, with Special Reference to Biological Variation".

Davies, G. R., "Introduction to Economic Statistics".

Davies, G. R., and Crowder, W. F., "Methods of Statistical Analysis".

Dawson, Shepherd, "An Introduction to the Computation of Statistics".

Day, E. E., "Statistical Analysis".

Edgeworth, F. Y., "Methods of Statistics".

Elderton, W. P., "Frequency Curves and Correlation".

Elderton, W. P., and Ethel M., "Primer of Statistics". (An introduction into laboratory practice.)

Ezekiel, M. J. B., "Methods of Correlation Analysis".

Falk, I. S., "The Principles of Vital Statistics".

Fisher, Arne, "The Theory of Probability"; also, "An Elementary Treatise on Frequency Curves".

Fisher, Irving, "The Making of Index Numbers; a study of their varieties, tests and reliability". (Several editions available.)

Forsyth, C. H., "An Introduction to the Mathematical Analysis of Statistics".

Garrett, H. E., "Statistics in Psychology and Education".

Gavett, G. I., "A First Course in Statistical Method".

Giddings, F. H., "Scientific Study of Human Society".

Giffen, Robert, "Statistics".

Gregory, C. A. and Renfrow, O. W., "Statistical Method in Education and Psychology" (Several editions).

Griffin, F. L. "An Introduction to Mathematical Analysis".

Haskell, A. C., "Graphic Charts in Business".

Hoffman, F. L., "Insurance Science and Economics". (Note Chapter 9 on application of averages.)

Jerome, Harry, "Statistical Method".

Jones, A. L., "Logic Inductive and Deductive: An Introduction to Scientific Method".

Jones, D. C., "A First Course in Statistics".

Jordan, D. F., "Business Forecasting".

Karsten, Karl G., "Charts and Graphs".

Kelly, Truman L., "Statistical Method".

Kent, F. C., "Elements of Statistics".

Keynes, J. M., "A Treatise on Probability".

King, W. I., "The Elements of Statistical Method".

Koren, John (Editor), "The History of Statistics; Their Development and Progress in Many Countries".

Lipka, Joseph, "Graphical and Mechanical Computation".

Lovitt, W. V. and Holtzclaw, H. F., "Statistics".

Marshall, W. C., "Graphical Methods".

Mayo-Smith Richmond, "Statistics and Sociology".

Meitzen, August, "History, Theory and Technique of Statistics".

Merriman, M., "A Textbook on the Method of Least Squares".

Mitchell, Wesley C., Bulletin 284, Bureau of Labor Statistics, "Index Numbers of Wholesale Prices in the United States and Foreign Countries, 1921" (Revision of Bulletin 173).

Mitchell, Wesley C., "Business Cycles".

Moore, Henry L., "Forecasting the Yield and Price of Cotton", "Economic Cycles: Their Law and Cause" and "Generating Economic Cycles".

Newsholme, Sir Arthur, "The Elements of Vital Statistics".

Pearl, Raymond, "Medical Biometry and Statistics".

Pearson, Karl, "Tables for Statisticians and Biometricians".

Pearson, Karl, "The Grammar of Science, Physical".

Riegel, Robert, "Elements of Business Statistics".

Rietz, H. L., ed., "Handbook of Mathematical Statistics".

Rietz, H. L., monograph on "Mathematical Statistics".

Rugg, H. O., "Statistical Methods Applied to Education".

Secrist, Horace, "An Introduction to Statistical Methods".

Secrist, Horace, "Readings and Problems in Statistical Methods".

Slichter, C. S., "Elementary Mathematical Analysis" (Various editions).

Social science research council. Advisory committee on social and economic research in agriculture, "Research Method and Procedure in Agricultural Economics" (2 volumes).

Thorndike, E. L., "An Introduction to the Theory of Mental and Social Measurements" (Various editions).

Tippett, L. H. C., "The Methods of Statistics" (An introduction mainly for workers in the biological sciences).

Wallace, H. A. and Snedecor, G. W., "Correlation and Machine Calculation". Ames, Iowa, Iowa State College of Agriculture and Mechanic Arts.

Walsh, C. M., "The Measurement of General Exchange Value".

Weld, L. D., "Theory of Errors and Least Squares". (Excellent introduction to measurement and properties of errors, with examples.)

West, Carl J., "Introduction to Mathematical Statistics".

Whipple, G. C., "Vital Statistics".

Whitaker, E. T. and Robinson, G., "The Calculus of Observations".

Yule, G. U., "An Introduction to the Theory of Statistics". (Several editions available. Excellent bibliography at the end of each chapter.)

Zizek, Franz, "Statistical Averages" (Translated by Warren M. Persons).

THE COLLECTION OF STATISTICAL DATA

I. Predetermined plan of procedure. (Jerome, pages 290-292; Secrist, revised edition, pages 47-53; Chaddock, pages 371-374).

a. Purpose of the investigation.

b. Determination whether the problem as formulated lends itself to statistical treatment.

c. Examination of available literature and previously collected data. (Secondary data).

1. Types of data necessary for the problem, analysis or solution.

2. Are the data available in suitable form?

3. Are the data adequate for the problem?

4. Have the data the required degree of accuracy, consistency, and comparability?

5. Proportion of problem which can be solved with secondary data; with primary data. Consideration must be given to time and money.

d. Choice of methods to be used in obtaining facts from secondary sources; from primary sources.

e. Selection of working force to collect secondary data; to collect primary data.

f. Preparation of the necessary forms and instructions for their use in the field by the working force.

g. Instructions for editing the schedules after they are returned by the enumerator or other informant to the statistician.

II. Methods of collecting data. (Note references below.)

a. Library method.

b. Field method.

c. Office or correspondence method.

1. Compulsory.

2. Cooperative.

III. Library method.

a. Definition: The library method involves the use of secondary sources, of compilations of facts prepared by some previous investigator. (Jerome, pages 292-294; Chaddock, pages 392-394; Crum and Patton, Chapter III; Day, Appendix B, pages 389-392).

b. Tests to be applied to secondary data before they are used. (These tests listed or discussed in most texts.)

1. Are the data homogeneous throughout the period for which they are available or the period for which they are to be utilized?

2. Is the validity of the data affected by an intentional or an unintentional bias?

3. Are there other elements of inaccuracy, due, for example, to the imperfection of the sources or of the methods used in collecting the data?

4. To summarize, are the data as adequate for the problem in hand as data which might be obtained from primary sources?

IV. Field method. (Jerome, pages 294-299; Chapin, "Field Work and Social Research", Chap. V.)

a. Source of data.

1. Official records: birth, death, and marriage certificate records.

2. Private records: private accounts, cost-of-living budgets, etc.

3. Individual unrecorded knowledge. (U.S.D.A. Bulletin 1020).

4. Direct observation.

b. Personnel of the investigating or working force.

1. Personal character of workers.

2. Standards for efficiency.

A. Study of conditions under which inquiry is to be made.

B. Definition of the statistical units.

I. Appropriateness for purpose of study.

II. Clarity.

III. Measurability.

IV. Comparability.

3. Forms to be used.

c. Scope of inquiry.

1. Complete.

2. Partial.

V. Office or correspondence method. (Jerome, pages 299-305; Chaddock, pages 374-395; Day, pages 383-388; Crum and Patton, Chapter IV; Lovitt and Holtzclaw, Chap. II).

a. Definition: The office method involves the substitution of inquiry by mail or by other means for field investigation.

b. Types of questionnaire studies.

1. Questionnaire - schedules filled in by a reporter.

2. Questionnaire - schedules filled in by an informant.

3. Questionnaire - schedules mailed to and filled in by an informant and later checked by an enumerator.

c. Questionnaire schedule filled in by a reporter.

1. Success of this method depends upon:

A. Personal ability of the individual making the interview.

B. Personal qualities of interviewer -- tact, diplomacy, courage, intellectual curiosity, integrity, etc.

C. Degree to which he understands:

I. The problem upon which he desires information.

II. The reactions of those he interviews.

D. Accuracy with which he:

I. Interprets information supplied.

II. Records and remembers facts submitted.

E. Form of record upon which answers are put.

d. Questionnaire schedule filled in by an informant.

1. Success of this method depends upon:

A. Locating and addressing questionnaire to those who have access to the facts.

B. Definition and precise statement of what is wanted.

C. Data requested should be as far as possible a matter of record rather than of opinion.

D. Formulating the questionnaire in such a manner that the units in which the data are measured are:

I. Same as those in current use.

II. Simple rather than being composite or expressed as ratios.

E. Ability of questionnaire to overcome indifference and reluctance of the informant to give information which may be:

I. Confidential.

II. Difficult or costly to assemble.

III. Of value to competitors.

c. Form and nature of questionnaire.

1. Explanatory letter should accompany the questionnaire.

A. The letter should state the purpose for which information is being collected.

B. It should quote the authority or auspices under which investigation is being made. Assurance should be given that inquiries are made according to the provisions of the law, or if voluntarily taken, with hope of throwing light on some problem.

C. It should give some inducement for an immediate reply.

D. It should promise a copy of results of the investigation to the informant.

E. Promise should be made to the informant that no individual data will be disclosed -- only totals and averages will be published.

2. Questionnaire should be accompanied by an addressed and stamped (or franked) envelope.

3. Form of questionnaire.

A. If possible the questionnaire should be of a convenient size to carry and send without folding.

B. If possible it should conform in size to some one of the standard filing devices, i.e., 5 x 8 inches, 8½ x 11 inches.

C. Quality of paper should be such as to permit use of ink, sufficiently durable to stand handling and sorting without replacement.

D. Careful attention should be given to the spacing and kind of type used in printing.

E. Number or letter each question or item in schedule to render easy identification and facilitate tabulation.

F. Experiment with sample questionnaires in a restricted locality to determine coverage, omissions, errors, etc.

G. If possible, it is desirable to furnish a duplicate for the records of the informant.

4. Form and nature of questions.

A. Units of measurement should be clearly indicated, accurately defined, and should conform with common usage. Units may appear at beginning, in body, or at end of questionnaire. For interview questionnaire, at beginning; for mail questionnaire, in proximity to entries to which they relate. Explanation of units should include a statement of the degree of numerical accuracy expected in answers.

B. The general appearance of questionnaire should not be crowded. Allow sufficient space for completion. Related questions should be grouped together.

C. False or inaccurate answers should be guarded against by having questions, so far as possible, corroboratory.

D. Questions should be simple, definite, clear, courteous, diplomatic, easy to answer, not inquisitorial. Inquirer should prevent evasive or double interpretation type of questions. Use "yes" or "no", "number" or "amount", or "check" questions. Use few questions; the fewer the questions, the greater the percentage return.

E. As a rule, the making of arithmetical calculations such as totals, ratios, percentages, etc., should be reserved for the statistical organization.

F. Rulings and columner arrangement should be simple and definite to prevent misplacing of items.

G. In every possible way reduce the task of the informant to a minimum in order to receive maximum returns. Ask informant to sign the questionnaire. This tends to increase the accuracy of his report.

f. Distribution of mail questionnaires.

1. To whom shall questionnaires be sent.

A. Method of sampling to be used.

2. All questionnaires to be sent out at the same time.

3. Determine closing date after which questionnaires will not be used. This aids in tabulating.

g. Return of mail questionnaires.

1. Success will probably attend project if

A. Informant is impressed with necessity of his completing and returning the questionnaire.

B. Mandatory power possessed.

C. Degree of cooperation between informant and inquirer is fairly high.

h. Editing of mail questionnaires.

1. Necessity of checking, editing and revision.

A. Editing is reduced to a minimum if questionnaire is properly made out.

2. Purpose of editing is to check:

A. Accuracy.

B. Consistency.

C. Uniformity of expression.

D. Completeness.

3. Editing.

A. As a rule the editor accepts as final the replies entered upon the schedule, but if there is a presumption of error he attempts to verify the answer.

B. Errors should be noted in a distinctive manner or color on the questionnaire.

C. Errors of omission, additions, false entry, and confusion of items may be readily corrected by editor. Alterations should be made only in case of an unmistakable error.

D. The degree to which omissions may be tolerated should be determined before beginning of editing. Study remainder of schedule and determine whether it is a representative sample.

i. Editing sheet.

1. Function.

A. To classify, arrange, and summarize in easily accessible form the answers to the questions with which the inquiry is concerned.

2. Method.

A. Preliminary: Analyze the factors in the problem investigated. Intelligent tabulating begins with making editing sheet such as to economically utilize results of inquiry.

B. Process of tabulation.

I. Form of editing sheet. Usually columnar in type divided into spacing with proper headings, rows numbered identically with questionnaire number, space for computing percentages, ratios, etc.

II. Tabulation in table.

III. Mechanical tabulations.

a. Various devices.

(1) Punch card.

(2) Key punch.

(3) Sorting machine.

(4) Total and listing machine.

j. Necessary budget for financing of investigation.

References

Bailey, W. B., and Cummings, John, "Statistics", Chap. III, pages 8-16, Chap. IV, pages 17-25.

Chapin, F. S., "Field Work and Social Research", Chap. VII, pages 148-191.

Rugg, H. O., "Statistical Methods Applied to Education", Chap. II, pages 39-56.

Secrist, H., "An Introduction to Statistical Methods", Revised Edition, Chap. II, III, IV, and V, pages 22-123.

Jerome, H., "Statistical Method", Chap. XVI and XVII, pages 290-336; also Chap. 2 (The sampling process).

Day, E. E., "Statistical Analysis", Appendix A and B, pages 383-392.

Chaddock, R. E., "Principles and Methods of Statistics", Chap. XIV, pages 371-396.

Pearl, Raymond, "Medical Biometry and Statistics", Chap. III.

Riegel, R., "Elements of Business Statistics", Chap. VI and VII.

Bowley, A. L., "An Elementary Manual of Statistics", part I, Chap. 7 and 8.

Crum and Patton, "Economic Statistics", Chap. III and IV, pages 22-33.

Lovitt and Holtzclaw, "Statistics", Chap. II and III.

F.J.H.
2.9.34

TABLE CONSTRUCTION

I. Recording of the data in table form is the next logical step after the sorting and counting are completed.

II. Definition of "table".
  a. A table is a triumph of ingenuity and technique, a masterpiece of economy of space combined with a maximum of clearly presented information. (Jerome, page 28)

III. Types of tables.
  a. General-purpose (primary) table.
    1. The general-purpose table is designed to bring together in the most convenient and accessible form all the data bearing upon a given topic. (Day, pages 404-405).
    2. The general-purpose table is designed mainly as a source of fundamental data, giving summations of individual items taken from original schedules and classified under a variety of categories. (Chaddock, page 404).
    3. The general-purpose table is a repository of facts compiled on some general topic for a particular time or period of time. (Sutcliffe, "Elementary Statistical Method, page 37).
  b. Special-purpose (secondary) table.
    1. A special-purpose table is intended to throw into relief relationships of special significance in a given study. It is a record of the results of a statistical analysis. (Day, page 405).
    2. Special-purpose tables are usually compiled from the data contained in the general purpose table and are arranged to emphasize the analysis of the results of a special study. (Sutcliffe, pages 38-39).
  c. Complete cross-classification table.
    1. It is one which indicates for each subclass of any one of the two or more classifications, the apportionment of the given subclass on the basis of any one of the other classifications appearing in the table. (Jerome, page 32).
  d. Partial cross-classification table.
  e. Partially subclassified table.

IV. Structural elements of a Table.
  a. Number.
  b. Title.
  c. Lettering.
  d. Ruling.
  e. Stub.
  f. Caption.
  g. Contents.
  h. Source and footnotes.

V. Principles of table construction. (Day, pages 403-404). (Also, see Mills, pages 73-74).

1. Every table should be self-sufficing, containing within itself a clear explanation of the meaning of the items displayed.
2. Every table should be logically a unit containing only data which are intimately related to one another.
3. Column-and row-headings should be brief, unambiguous, and self-explanatory, table footnotes being used when necessary to make the headings perfectly clear.
4. Coordinate and subordinate relationships among the column- and row-headings should be shown by variations of boxing in the captions and of indentation in the stub.
5. Varieties of letters, figures, lines, column-widths, and inter-linear spacings should be employed to facilitate easy and intelligent use of the table.
6. Columns and rows should be lettered or numbered if cross-reference is desirable.
7. Sources and units should invariably be indicated.

Statistical Tables

A statistical table is an arrangement of data in rows and columns according to some definite classification. Classification is the combination of similar or like data into groups. The purpose of a table is to arrange data in a condensed form to facilitate comparisons. A table consists of (1) number, (2) title, (3) source, (4) legend, (5) rulings, (6) stubs, (7) captions, and (8) contents.

(1) The purpose of the number is to make reference to the table simple and clear. In a large table columns and rows should be numbered or lettered also.

(2) The title should be clear, concise and catchy. It should tell exactly what is in the table, when the data were gathered, or events occurred, where data were collected from, and if possible, the classifications made. If title is too long use a main title and a sub-title with different sized lettering to emphasize important parts.

(3) The source of the data should be given. If you are the original collector of the data, it should be noted in the article in which the table appears rather than in a footnote. Otherwise, a footnote giving the source should accompany each table.

(4) The legend should tell in what units the data in the table are expressed. If all of the data in the table are expressed in the same unit, the legend may be placed above the top ruling or directly under the title. If different units are used for different stubs or captions, the legend should appear at the head of the column or at the left of the row.

(5) Each columnar line and each space between rows in the table should have a definite purpose. There are several styles of ruling tables, but one should decide upon the style best suited for this purpose and follow it rigidly.

(6) and (7) Stubs and captions should be arranged in some definite order, i.e. chronological, alphabetical, geographical, or according to some natural division of the subject matter. The stubs and captions should adhere to the definitions or limitations expressed by the stubs and captions, respectively, and if repeated should always be arranged in the same order each time. They should be brief, but long enough to be self-explanatory. They must be rigidly defined.

(8) Only data that are closely related should be shown in the same table. All possible repetition of figures in a table should be avoided. Every table should be as simple as possible. Simplification leads to a better understanding of the contents and brings out more clearly the comparisons that are intended. The data should be arranged so that comparisons to be most frequently made can be most easily made.

S A M P L E

Table 1.--Classification of second hand fords at A. Bee Ford Agency, January 1, 1930, showing the type, model, year of car and kind of wheels

Type of car and year of make	Model of car and kind of wheels								
	Total cars			Wire wheels			Wooden wheels		
	Total	Model A	Model T	Total	Model A	Model T	Total	Model A	Model T
	No.	No.	No.	No.	No.	No.	No.	No.	No.
Total	163	88	75	104	78	26	59	10	49
Roadster	25	16	9	16	15	1	9	1	8
Touring	13	5	8	9	5	4	4	0	4
Coupe	48	27	21	32	25	7	16	2	14
Tudor	44	22	22	28	21	7	16	1	15
Fordor	33	18	15	19	12	7	14	6	8
1928 model total	118	65	53	79	57	22	39	8	31
Roadster	18	13	5	13	12	1	5	1	4
Touring	9	4	5	7	4	3	2	0	2
Coupe	35	20	15	24	18	6	11	2	9
Tudor	34	18	16	22	17	5	12	1	11
Fordor	22	10	12	13	6	7	9	4	5
1929 model total	45	23	22	25	21	4	20	2	18
Roadster	7	3	4	3	3	0	4	0	4
Touring	4	2	2	2	1	1	2	0	2
Coupe	13	8	5	8	7	1	5	0	5
Tudor	10	6	4	6	4	2	4	0	4
Fordor	11	6	5	6	6	0	5	2	3

John Doe Statistical Agency, New Cork, New Jersey. Compiled from records of the Automobile Market Corporation, Old City, New York.

Note: Order of importance: (1) type of car, (2) model of car, (3) year of make, and (4) kind of wheels. Attention is called to the placement of totals.

FJH
2-9-34

GRAPHIC PRESENTATION

I. Purpose of Graphics.
   a. For the simplification, visualization, interpolation, and interpretation of quantitative data.

II. Standards of Diagrammatic Presentation. (Secrist, pages 212-213; Jerome, pages 51-52; Sutcliffe, page 52; Mills, pages 51-59).
   a. The statistical facts and graphic device must agree.
   b. The graphic device must be of an appropriate form selected according to psychological appeal and ease of comprehension.
   c. The illustrations used should be those least misunderstood, and most faithfully and correctly interpreting the facts.
   d. Charts should be verifiable by reader. Indicate on the diagrams the scales or values used. Include as a part of the chart the data which the chart illustrates.
   e. Graphic devices are to be considered as illustrations of analyses rather than methods by which analyses are made.

III. Graphic Forms: Time Element Constant. (First numbers refer to chart numbers in Jerome; the second, to Mills, Blank indicates no illustration).
   a. One variable.
      1. Horizontal bar chart.
         A. Simple horizontal bar chart. (1; -)
         B. Sub-divided horizontal bar chart. (3; -)
         C. Percentage horizontal bar chart. (4; -)
      2. Statistical maps.
         A. Cross-hatch map. (8; -)
         B. Multiple dot map. (9;-)
         C. Quartered dot map. (10; -)
         D. Graduated dot map. (-;-)
      3. Area and volume charts.
         A. Pie diagram.
            I. Open-face type. (11;-)
            II. Cross-hatch type. (12;-)
         B. Emblematic pictograms.
      4. Frequency charts.
         A. Discrete series.
            I. Vertical bar chart. (7;-)
         B. Continuous series.
            I. Histogram. (20 and 21; 22, 23, 24, 25, 29, 30, and 31)
            II. Polygon. (20 and 21; 26, 27, 28, 40, 41, and 42)
            III. Smoothed curve. (20 and 21; 32, 36, and 43).
            IV. Cumulative (ogive) curve. (24; 33, 34, and 35).

b. Two variables.
1. Scatter diagram. (39; 67, 72, 80)
2. Double frequency diagram. (-; 68, 79)

IV. Graphic Forms: Time Element Varying
a. One variable.
1. Arithmetic scales.
A. Horizontal bar chart. (2;-)
B. Vertical bar chart. (5;-)
C. Time (line) graph or historigram. (13; 13, 15, 64, 74, 75)
I. Simple.
II. Cumulative.
D. Special graphs.
I. Component-parts band chart. (14;18)
II. Silhouette excess-and-deficit chart. (15;-)
III. Year-month chart. (16;-)
IV. Gantt progress chart. (-;21)
2. Log-arithmetic scales.
A. Ratio chart (22 and 23; 12, 14, and 16)
3. Log-log scales. (-;83)
b. Two variables.
1. Scatter diagram. (-;81, 82, and 84).
2. Double frequency scatter diagram.

V. Working rules for chart making. (See Jerome, pages 52-58; 75-78, pages 104-106; Mills, pages 51-59).

SUMMARY DEVICES

I. Methods of Describing a Frequency Distribution.
   a. Graphic devices.
      1. Array.
      2. Frequency distribution.
         A. Histogram.
         B. Polygon.
         C. Smoothed curve.
   b. Disadvantages of graphic devices.
      1. Difficulty of analysis and comparison because of different scale units.
      2. Description of graphic devices or comparison of graphic devices is possible only in descriptive or comparative words or approximations which are inaccurate.
      3. Because the human mind cannot ably grasp and retain in proper relation curve comparisons, an analytical comparison of two or more graphic devices simultaneously is grossly difficult and usually inaccurate.
   c. Types of analytical summary devices necessary.
      1. A measure of central tendency, i. e., an average.
      2. A measure of the degree of dispersion and concentration about the central tendency, i. e., a measure of variation or dispersion.
      3. A measure of the lack of symmetry; i. e., skewness.
      4. A measure of the degree of peakedness of the frequency distribution, i. e., kurtosis.

F. J. H.
2/9/34

FREQUENCY DISTRIBUTION

I. Classification of Quantitative Data.
  (a) Purpose.
    (1) To bring order out of confusion of many individual values.
    (2) To permit a preliminary analysis of the data.
  (b) Methods.
    (1) Array: rank of values in order of size.
      A. Advantages.
        1. Permits calculation of range.
        2. Ability to note point of central tendency.
        3. The characteristics of data begin to appear.
    (2) Frequency distribution.
      A. Magnitude classes or class intervals.
        1. Advantages.
          a. It is of a simpler form than an array.
          b. More accurate indication of regularities of distribution.
          c. Comparison with other distributions is possible from standpoints of central tendencies, ranges, dispersion, kurtosis, etc.
      B. Assumptions underlying the formation of frequency classes.
        1. Even or uniform distribution of items in each class interval, or
        2. Concentration of all items at the mid-value of the class, or
        3. Distribution of items in each class interval in form of normal curve.
      C. Size of class interval (or number of classes).
        1. Size of class interval should be adjusted without material sacrifice of accuracy.
          a. Arrange classes so there will be an even distribution throughout each class, or the average of all values in each class will equal the mid-value.
          b. The number of classes should be so determined that an orderly and regular sequence of frequencies is secured.
          c. Number of classes should be limited to 15 to 20 in order that data can be manipulated easily and their significance grasped and appreciated.
          d. In data that cluster about a point, make the point if possible a mid-point of an interval.
          e. Class intervals should be equal in width.
          f. Size of class intervals in a continuous series should be taken to the point at which the histogram represents clearly the distribution of an infinite number of samples.

II. Graphic Presentation.
(a) An array.
(b) Frequency distribution.
(1) Histogram.
(2) Polygon.
(3) Smoothed Curve.

III. Histogram.

IV. Polygon.

V. Smoothed Curve.
(a) Purpose.
(1) Estimate the distribution of frequencies within a given class.
(2) To smooth out accidental irregularities shown by the sample and thus more nearly to approximate the graph which would best represent the universe (population) from which the sample was taken.
(b) Method.
(1) Connect mid-point of each class interval by a line (polygon).
(2) The curve should be so smoothed that the total area under the resulting curve is equal to the sum of the area of the original rectangles.
(3) Where possible, the areas of the individual rectangles are to remain approximately unchanged.
(4) The curve must be free from abrupt changes in direction.
(5) Ordinarily the peak of smoothed curve should equal or slightly over-top the peak of the frequency polygon.
(6) In general, the nearer the approach to a truly continuous series, the more appropriate does the smoothing process become.

ELEMENTARY STATISTICS

2nd Semester

I

Draw or sketch roughly scatter diagrams to show

(a) No correlation
(b) High negative correlation
(c) High non-linear correlation
(d) Low positive correlation

In each scatter diagram use "Y" as the dependent variable and "X" as the independent variable. Show the mean of "X" (M_x) and the mean of "Y" (M_y) on each diagram.

II

"The regression equation constitutes a measure of the functional relationship between a dependent variable and an independent variable or variables, but the equation is only an expression of average relationship."

(a) What average is meant by the word, "average"? Explain briefly.

(b) Name the statistical device and give its formula which will measure the variation or dispersion of the data about the regression equation.

(c) State in one or two sentences why the statistical measure indicated in (b) above is not a satisfactory measure of correlation.

(d) Show:

1. That when $r = o$, $S_y = \sigma_y$

2. The condition or conditions under which "r" may be equal to 1 (plus or minus).

III

Solve one of the following correlation problems:

(a) $\Sigma fd_y = 200$ Assumed mean of Y = 5
$\Sigma fd_x = 50$ Assumed mean of X = 6
$\Sigma fd_y^2 = 1300$ Class interval of Y = 2
$\Sigma fd_x^2 = 1625$ Class interval of X = 8
$\Sigma fd_x d_y = 900$
$n = 100$

Compute "r_{yx}", "a", and "b_{yx}" of the equation, $Y = a+bX$, and "S_y". Assume "Y" is the dependent variable and "X" is the independent variable.

III (Continued)

(b) $\Sigma XY = 25{,}000$ $\quad n = 100$
$\Sigma X^2 = 60{,}000$ $\quad M_y = 10$
$\Sigma Y^2 = 20{,}000$ $\quad M_x = 20$

Using "Y" as the dependent variable and "X" as the independent solve for "b_{yx}", "a", "r_{yx}", and "S_y".

(c) $\Sigma fd_z = 10$ $\quad \Sigma fd^2_x = 1625$
$\Sigma fd_z^2 = 170$ $\quad$ Class interval of Y = 2
$\Sigma fd_y = 200$ $\quad$ Class interval of X = 8
$\Sigma fd_x = 50$ $\quad n = 10$

$\Sigma fd^2_y = 1300$

Assuming "Y" is the dependent variable and "X" is the independent, calculate "r_{yx}", "S_y" and "b_{yx}".

IV

(a) Answer each of the following questions using only one short sentence for each question:

1. When may r_{yx} be equal to f_{yx} ?
2. When may r_{12} be equal to $r_{12.3}$?
3. When may r_{12} be equal to $R_{1.23}$?
4. When may $R^2_{1.234}$ be equal to $d_{12.34}$?
5. When may b_{yx} be equal to zero?

(b) Write the normal equations for a multiple correlation problem in which X_2 is the independent variable, X_1, X_3, and X_4 are independent variables. Write the normal equations in terms of "p's" and "σ's".

V

(a) What conclusion or conclusions would you draw from these results: (Be brief)

1. $r_{12} = .60$; $\rho_{12} = .60$; $R_{1.23} = .60$; $R_{1.24} = .80$

2. $\sigma_1 = 4$; $S_{1.2} = 3$; $S_{1.23} = 2$; $S_{1.234} = 1$

3. $r_{12} = .80$; $r_{12.3} = .40$; $r_{12.345} = .0001$

4. $r_{12} = +.90$; $r_{12.34} = -.90$

(b) Explain briefly the difference between:

1. $b_{12.34}$ and $\beta_{12.34}$

2. σ_y and S_y

3. $d_{12.34}$ and $\beta_{12.34}$

QUESTIONS ON THE COLLECTION OF STATISTICAL DATA

1. Outline briefly the procedure necessary to carry on a statistical investigation.

2. Enumerate the factors or elements which determine whether a census is a good one or not.

3. (a) An agricultural experiment station is interested in the marketing of tomatoes. The tomato growers contend that the use of fertilizer rich in potash tends to toughen and strengthen the skin of the tomato so that it may be shipped long distances. You have been chosen to plan an investigation which will verify or contest the opinion of the growers. Explain BRIEFLY the steps necessary in the preparation for the actual work of investigation.

   (b) Indicate briefly what method of procedure would be most appropriate and practical for an investigation into:

      1. Buying habits of women customers.

      2. Causes of juvenile delinquency.

      3. Changes in the cost of living in Washington, D. C. from 1933 to 1934.

      4. A statistical measurement of recreational facilities in a rural, sub-urban, and urban community.

      5. Relation between rent received and land values.

   (c) Discuss briefly the problems you would expect to encounter in constructing a series of monthly average prices for a particular commodity in a particular market from actual daily sales in that market.

   (d) To what sort of problems are statistics and statistical methods inappropriate?

4. (a) What tests whould a statistician apply to secondary data before using them?

   (b) Name at least ten general sources of secondary statistical data.

5. Name and explain the advantages and disadvantages of the various methods of collecting data from original sources.

6. Explain the outstanding characteristics of a well-prepared questionnaire.

7. Criticise the following questions found on questionnaires:

   (a) "Did you read this magazine from cover to cover?"

   (b) "What suggestions do you make for development of college spirit at your institution?"

   (c) "How much cotton, corn, wheat, oats, etc., have you sold in the past year?"

   (d) "How large is your farm?"

   (e) "How many members of your family have attended college?"

   (f) "What is the trend of the death rate in your community?"

   (g) "Name the biggest man in your city".

   (h) "What is the rate of labor turnover in your factory?"

8. Formulate a good usable definition of the following statistical units:

   (a) A "residence", for use in studying housing conditions in Washington, D. C.

   (b) An "Improved farm", to be used in a farm survey in Maryland.

   (c) A "family", to be used in a cost-of-living analysis.

   (d) A "retail merchant", for use in a distribution study.

   (e) An "insane person".

   (f) An "establishment", for use in a census of manufactures.

   (g) A "criminal".

9. (a) Distinguish briefly between the library, field and correspondence methods of obtaining data.

   (b) "Data which are primary in the hands of one statistician may be secondary in the hands of another". Explain and give examples.

10. "Editing of schedules has in mind four purposes". Explain.

11. Name the requisites of a good tabulation sheet or schedule.

Required Materials for "Elementary Statistics"

1. Text:

   (a) Mills, F. C., "Statistical Methods Applied to Economics and Business". This book may be secured from several sources, including Washington book stores (second-hand or new book stores), from friends, or by ordering a new copy through the instructor, who in turn will order it through the Bureau of Agricultural Economics Library.

2. Suggested Supplies: (may be secured at a book or stationery store)

   (a) A heavy covered notebook in which to file mimeographed material and take notes on lectures and on material contained in the textbook or other books.

   (b) Pad of graph paper (10 spaces to an inch, preferably blue in color).

   (c) Five sheets of 2 cycle semi-log paper.

   (d) Fountain pen, pencil, ruler, ink and pencil eraser, and art gum.

Class Regulations

3. (a) The class meets twice each week from 4:50 to 5:50 P.M. Roll will be called, and attendance, non-attendance or tardiness will be recorded.

   (b) If a student is unable to attend a class, the assignment may be obtained by phoning Mr. B. R. Stauber, Extension 2065, or Mr. F. J. Hosking, Extension 2241. Both phones are on the District 6350.

   (c) All assigned problems must be completed before a grade will be issued for the course. Each individual piece of work will receive a full grade when it is completed and submitted by the

designated time, but a deduction of one letter will be made for each week an assigned problem is late. A reasonable excuse will be accepted if work cannot be handed in on time.

(d) The following letters will be used to indicate the quality of work submitted:

A - Excellent
B - Good
C - Fair
D - Pass
E - Condition
F - Failure

(e) 5, 10, or 15 minute examinations will be given from time to time without previous announcement. An examination of an hour or more in length will be given at mid-semester and/or at the close of the semester.

(f) Grades on the assigned problems make up about one half of the final grade, recitations and short-time quizzes about one fourth, and the mid-semester and final examinations about one fourth.

1. NATURE AND USES OF STATISTICAL METHOD

I. Meaning of the word, "statistics".

a. Origin of the word, "statistics".

1. "Statistics" appears to be derived from the Italian, "ragione di stato", practical politics, and "statista", statesman.

b. Historical development of the word, "statistics".

1. "Statistics is the science that teaches us what is the political arrangement of all the modern states of the known world". This is considered one of the earliest known definitions of "statistics". (Noted in "The Elements of Universal Erudition", J. F. Bielfeld, 1770. From Yule, "An Introduction to the Theory of Statistics", page 1.)

2. A wider definition appears in the preface to "A Political Survey of the Present State of Europe", E. A. W. Zimmerman, 1787. He states that "that branch of political knowledge which has for its object the actual and relative power of the several modern states, the power arising from their natural advantages, the industry and civilization of their inhabitants, and the wisdom of their governments, has been formed.....into a separate science". (From Yule, page 1.)

3. The early exposition of the noteworthy characteristics of the state, or "statistics", was preponderantly verbal. In time "statistics" acquired a more narrow meaning, viz., the exposition of the characteristics of a state by numerical methods.

4. The word "statistics" was transferred to those data with which it operated, such as vital statistics, economic statistics, etc.

5. Today, the term statistics is used somewhat indiscriminately to refer to any one of three concepts - statistics, statistical method, and statistical theory.

II. Explanation of the modern concepts of "statistics".

a. Statistics or statistical facts.

1. Quantitative data affected to a marked degree by a multiplicity of causes. (Yule, page 5.)

2. Aggregates of facts numerically expressed. (Secrist, "An Introduction to Statistical Method", page 10.)

b. Statistical method or technique.

1. The science of counting. (Bowley, "Elements of Statistics", 4th edition, page 3.)

2. The elucidation of quantitative data affected by a multiplicity of causes. (Yule, page 5.)

3. The term "statistics", when used to designate a branch of study implies an exposition of certain methods employed in presenting and interpreting the numerical aspects of a given subject. (Davies, "Introduction to Economic Statistics", page 3.)

4. The method of judging collective natural or social phenomena from the results obtained by the analysis of an enumeration or collection of estimates. (King, "Elements of Statistical Method", page 23.)

5. The principles of the collection and the analysis of numerical data and the accurate and effective presentation of such data and the results of their analysis. (Jerome, "Statistical Method", page 2.)

c. Statistical theory.

1. The exposition of statistical methods. (Yule, page 5.)

2. Mathematical and economic principles which underlie and justify statistical methods. (Jerome, page 1.)

III. The uses of statistical method.

a. Pre-eminent causes for modern importance of statistics.

1. As listed by Jerome, pages 3-5.

A. This is an age of large numbers.

B. Modern science demands its theories be based on facts.

C. Business is becoming more and more a matter of scientific procedure.

2. As listed by Day, "Statistical Analysis", pages 6-8.

A. Growing need of more precise measurement in human affairs.

B. Spread of the modern scientific spirit.

C. Statistical method is the only science able to study social complexes.

b. Application of statistical method.

1. As listed by Jerome, pages 5-7.

A. Analysis of social phenomena.

B. Economic conflicts.

C. Commercial and political use.

I. Public administration.

II. Internal and external business statistics.

2. As listed by Chaddock, "Principles and Methods of Statistics", pages 31-39.

A. Statistics in the service of social sciences.

I. Dependence of the social sciences on statistics.

B. Service of statistics to economics and sociology.

C. Statistics in the service of education.

3. As listed by Mills, "Statistical Methods", chapter I.

c. Requisites for success in statistical study and practice.

1. Constructive imagination, intellectual integrity, capacity for careful work, and knowledge of subject matter. (Jerome, pages 9-11.)

2. Good judgment, broad knowledge and experience, and common sense are the most valued possessions of a statistician. (Chaddock, page 31.)

References

(These readings are suggested for the purpose of presenting different points of view from which writers have approached the study of statistical methods.)

Mills, F. C., "Statistical Methods", chapter 1.
King, W. I., "Elements of Statistical Method", part I.
Chaddock, R. E., "Principles and Methods of Statistics", part I, pages 3-39.
Pearson, Karl, "The Grammar of Science, Physical", part I, 3rd edition, chap. 1.
Yule, G. U., "An Introduction to the Theory of Statistics", Introduction.
Jerome, Harry, "Statistical Method", chapter 1.
Mayo-Smith, R., "Statistics and Sociology", chapters 1, 2, and 3.
Whipple, G. C., "Vital Statistics", 2nd edition, chapter 1.
Pearl, Raymond, "Medical Biometry and Statistics", chapter 1.
Jones, D. C., "A First Course in Statistics", chapter 1, page 4.
Burgess, R. W., "Introduction to the Mathematics of Statistics", chapter 1.
Lovitt, W. V., and Holzclaw, H. F., "Statistics", chapter 1.

QUESTIONS ON
THE NATURE AND USES OF STATISTICAL METHOD

1. Outline the method of forming a judgment of ideas and concepts. What proportion of this process is quantitative?

2. (a) Compare the development of the methods of statistics with the growth of nations.

   (b) How can statistics be used as an aid in the development of a powerful nation?

3. Criticize:

   (a) "Statistics may, for instance, be called the science of counting".

   (b) "The term 'statistics' when used to designate a branch of study implies an exposition of certain methods employed in presenting and interpreting the numerical aspects of a given subject".

4. Distinguish between statistical facts, statistical theory and statistical methods. With which are we dealing in this course?

5. Distinguish between accounting and statistics.

6. Name and explain the reasons for the modern importance of statistics to students of agriculture and economics.

7. What evidence is there to support the contention that there is an increasing need for the development of statistical science in agriculture?

8. What concrete illustrations can be given of the use of statistical method in agriculture? In business? In sociological study? In the administration of the laws of economic significance? In labor conflicts?

9. What relation is there between the statement that intellectual integrity is an essential for statistical practice and the frequently heard assertion that "anything can be proved with figures"?

10. What personal qualities are requisite for success as an agricultural statistician? Are the same qualities necessary in a student of statistics? Why?

Materials for "Elementary Statistics"

1. Text:

   (a) Mills, F. C., "Statistical Methods". This book may be secured from Washington book stores (second-hand or new book stores), from friends or acquaintances, or from the instructor, who will order it through the Bureau of Agricultural Economics Library. The supply of second-hand books is very small. Other statistical books are listed in a mimeographed bibliography.

2. Suggested Supplies: (may be secured at a book or stationery store)

   (a) A heavy covered notebook in which to file mimeographed material, problems and notes on lectures and on material contained in the textbook or other books.

   (b) Pad of graph paper (10 spaces to an inch, preferably blue in color).

   (c) Five sheets of 2 cycle semi-log paper.

   (d) Fountain pen, pencil, ruler, ink and pencil eraser, and art gum.

Class Regulations

3. (a) The class meets twice each week at a time designated by the instructor. The class will meet for approximately one hour. Roll will be called, and attendance, non-attendance or tardiness will be recorded.

   (b) If a student is unable to attend a class, the assignment may be obtained by telephoning the instructor.

(c) All assigned problems must be completed satisfactorily before a grade will be given for the course. Each individual piece of work will receive a grade when it is completed and submitted by the designated time, but a deduction of one letter will be made if an assigned problem is late. The instructor will indicate at time of assignment when a problem is due. A reasonable excuse will be accepted if work cannot be handed in at the designated time.

(d) The following letters will be used to indicate the quality of work submitted:

A - Excellent
B - Good
C - Fair
D - Pass
E - Condition
F - Failure

A paper receiving "E" must be reworked and submitted.

(e) 5, 10, or 15 minute examinations will be given from time to time without previous announcement. An examination of an hour or more in length will be given at mid-semester and/or at the close of the semester.

(f) Grades on the assigned problems make up about one-half of the final grade, recitations and short-time quizzes about one-fourth, and the mid-semester and final examinations about one-fourth. These proportions, of course, may be varied by the instructor.

NATURE AND USES OF STATISTICAL METHOD

I. Meaning of the word, "statistics".

a. Origin of the word "statistics".

1. "The words 'statist', 'statistics', 'statistical' appear to be all derived, more or less indirectly from the Latin, 'status', in the sense that it acquired in medieval Latin of a political state". Yule, "An Introduction to the Theory of Statistics", Page 1.

b. Historical development of the word, "statistics".

1. "Statistics is the science that teaches us what is the political arrangement of all the modern states of the known world". This is considered one of the earliest known definitions of "statistics". (Noted in "The Elements of Universal Erudition", Baron J. F. von Bielfeld, 1770. From Yule, Page 1.

2. A wider definition appears in the prefact to "A Political Survey of the Present State of Europe", E. A. W. Zimmerman, 1787. He states that "that branch of political knowledge which has for its object the actual and relative power of the several modern states, the power arising from their natural advantages, the industry and civilization of their inhabitants, and the wisdom of their governments, has been formed.....into a separate science". As used by Zimmerman and his contemporaries, "statistics" meant simply the exposition of the noteworthy characteristics of a state, the mode of expression being preponderously verbal. (Yule, Pages 1 and 2).

3. In time "statistics" acquired a more narrow meaning, viz., the exposition of the characteristics of a state by numerical methods.

4. Gradually, the word "statistics" was transferred to those data with which it operated, such as vital statistics, economic statistics, etc.

5. Today, the term statistics is used somewhat indiscriminately to refer to any one of three concepts--statistics, statistical method, and statistical theory.

II. Explanation of the modern concepts of "statistics".

a. Definition of "statistics" or "statistical" facts.

1. Quantitative data affected to a marked extent by a multiplicity of causes. (Yule, page 5.)

2. Aggregates of facts numerically expressed. (Secrist, "An Introduction to Statistical Method", page 10.)

3. Statistics means....compiled data, being the systematic quantitative expression of facts. (Smith, J. G., "Elementary Statistics", Page 8).

b. Definition of "statistical" method or technique.

1. The science of counting. (Bowley, "Elements of statistics", 4th edition, page 3.)

2. The elucidation of quantitative data affected by a multiplicity of causes. (Yule, page 5.)

3. This branch of study implies an exposition of certain methods employed in presenting and interpreting the numerical aspects of a given subject. (Davies, "Introduction to Economic Statistics", page 3.)

4. The method of judging collective natural or social phenomena from the results obtained by the analysis of an enumeration or collection of estimates. (King, "Elements of Statistical Method", page 23.)

5. The principles of the collection and the analysis of numerical data and the accurate and effective presentation of such data and the results of their analysis. (Jerome, "Statistical Method", page 2.)

c. Statistical theory.

1. The exposition of statistical methods. (Yule, page 5.)

2. Mathematical and economic principles which underlie and justify statistical methods. (Jerome, Page 1).

3. Statistical theory means that body of principles which has been developed to serve as a guide for sound statistics and statistical methods (Smith, page 8).

III. Uses of statistical method.

a. Pre-eminent causes for modern importance of statistics.

1. As listed by Jerome, pages 3-5.

A. This is an age of large numbers.

B. Modern science demands its theories be based on facts.

C. Business is becoming more and more a matter of scientific procedure.

2. As listed by Day, "Statistical Analysis", pages 6-8.

A. Growing need of more precise measurement in human affairs.

B. Spread of the modern scientific spirit.

C. Statistical method is the only science able to study social complexes.

b. Application of statistical method.

1. As listed by Jerome, pages 5-7.

A. Analysis of social phenomena.

B. Economic conflicts.

C. Commercial and political use.

I. Public administration

II. Internal and external business statistics.

2. As listed by Chaddock, "Principles and Methods of Statistics, pages 31-39.

A. Statistics in the service of social sciences.

I. Dependence of the social sciences on statistics.

B. Service of statistics to economics and sociology.

C. Statistics in the service of education.

3. As listed by Mills, "Statistical Methods", chapter I.

c. Requisites for success in statistical study and practice.

1. Construction imagination, intellectual integrity, capacity for careful work, and knowledge of subject matter. (Jerome, pages 9-11.)

2. Good judgment, broad knowledge and experience, and common sense are the most valued possessions of a statistician. (Chaddock, page 31).

References

(These readings are suggested for the purpose of presenting different points of view from which writers have approached the study of statistical methods.)

Text:
Mills, F. C., "Statistical Methods", chapter 1.

Other Readings:
Burgess, R. W., "Introduction to the Mathematics of Statistics". chapter 1.
Chaddock, R. E., "Principles and Methods of Statistics", part I, pages 3-39.
Davies, G. R., and Crowder, W. F., "Methods of Statistical Analysis", chapter 1.
Jerome, Harry, "Statistical Method", chapter 1.
Jones, D. C., "A First Course in Statistics", chapter 1, page 4
King, W. I., "Elements of Statistical Method", part I.
Lovitt, W. V., and Holzclaw, H. F., "Statistics", chapter 1.
Mayo-Smith, R., "Statistics and Sociology", chapters 1, 2, and 3.
Pearl, Raymond, "Medical Biometry and Statistics", chapter 1.
Pearson, Karl, "The Grammar of Science, Physical", part I, 3rd edition, chap. 1.
Smith, J. G., "Elementary Statistics", chapter 1.
Whipple, G. C., "Vital Statistics", 2nd edition, chapter 1.
Yule, G. U., "As Introduction to the Theory of Statistics". Introduction.

QUESTIONS ON THE NATURE AND USES OF STATISTICAL METHOD

1. Outline the method of forming a judgment of ideas and concepts. What proportion of this process is quantitative?

2. (a) Compare the development of the methods of statistics with the growth of nations.
   (b) How can statistics be used as an aid in the development of a powerful nation?

3. Criticize:

   (a) "Statistics may, for instance, be called the science of counting".

   (b) "The term 'statistics' when used to designate a branch of study implies an exposition of certain methods employed in presenting and interpreting the numerical aspects of a given subject."

4. Distinguish between statistical facts, statistical theory and statistical methods. With which are we dealing in this course?

5. Distinguish between accounting and statistics.

6. Name and explain the reasons for the modern importance of statistics to students of agriculture and economics.

7. What evidence is there to support the contention that there is an increasing need for the development of statistical science in agriculture?

8. What concrete illustrations can be given of the use of statistical method in agriculture? In business? In sociological study? In the administration of the laws of economic significance? In labor conflicts?

9. What relation is there between the statement that intellectual integrity is an essential for statistical practice and the frequently heard assertion that "anything can be proved with figures"?

10. What personal qualities are requisite for success as an agricultural statistician? Are the same qualities necessary in a student of statistics? Why?

TEXT AND REFERENCES

Text: Mills, F. C., "Statistical Methods". Bibliography given on pages 589-594.

Suggested References:

Adams, Thomas S. (editor), "Manual of Charting".

American Telephone and Telegraph Co., Statistical Methods Series No. 1, issued by Chief Statistician; "Introduction to Frequency Curves and Averages"; "Statistical Analysis and Projection of Time Series"; also "Graphical Method of Smoothing a Series of Frequency Curves".

Bailey, W. B., and Cummings, John, "Statistics".

Barlow, "Tables of Squares, Cubes, Square Roots, Cube Roots, Reciprocals".

Bowley, A. L., "The Measurement of Social Phenomena".

Bowley, A. L., "Elements of Statistics"; also, "An Elementary Manual of Statistics" (Various editions).

Brinton, W. C., "Graphic Methods for Presenting Facts".

Broad, C. D., "On the Relation between Induction and Probability".

Brunt, David, "The Combination of Observations".

Burgess, R. W., "Introduction to the Mathematics of Statistics".

Chaddock, R.E., "Principles and Methods of Statistics".

Chambers, G. G., "An Introduction to Statistical Analysis".

Chapin, F. S., "Field-Work and Social Research".

Coolidge, J. L., "An Introduction to Mathematical Probability".

Connor, L. R., "Statistics in Theory and Practice".

Crum, W. L. and Patton, A. C., "An Introduction to the Methods of Economic Statistics".

Davenport, C. B., "Statistical Methods, with Special Reference to Biological Variation".

Davies, G. R., "Introduction to Economic Statistics".

Davies, G. R., and Crowder, W. F., "Methods of Statistical Analysis in the Social Sciences".

Dawson, Shepherd, "An Introduction to the Computation of Statistics".

Day, E. E., "Statistical Analysis".

Dittmer, Clarence G., "Introduction to Social Statistics".

Edgeworth, F. Y., "Methods of Statistics".

Elderton, W. P., "Frequency Curves and Correlation".

Elderton, W. P., and Ethel M., "Primer of Statistics". (An introduction to Laboratory Practice.)

Ezekiel, M. J. B., "Methods of Correlation Analysis".

Falk, I. S., "The Principles of Vital Statistics".

Fisher, Arne, "The Mathematical Theory of Probabilities"; also, "An Elementary Treatise on Frequency Curves".

Fisher, Irving, "The Making of Index Numbers; a study of their varieties, tests and reliability". (Several editions available.)

Fisher, R.A., "Statistical Methods for Research Workers".

Forsyth, C. H., "An Introduction to the Mathematical Analysis of Statistics".

Florence, P. Sargant, "The Statistical Method in Economics and Political Science".

Garrett, H. E., "Statistics in Psychology and Education".

Gavett, G. I., "A First Course in Statistical Method".

Giddings, Franklin H., "The Scientific Study of Human Society".

Giffen, Robert, "Statistics".

Glover, J. W., "Tables of Applied Mathematics in Finance, Insurance, Statistics".

Gregory, C. A. and Renfrow, O. W., "Statistical Method in Education and Psychology" (Several editions).

Griffin, F. L. "An Introduction to Mathematical Analysis".

Haney, L. H., "Business Forecasting".

Hardy, C. O., and Cox, G. V., "Forecasting Business Conditions".

Harper, F. H., "Elements of Practical Statistics".

Haskell, A. C., "Graphic Charts in Business".

Hoffman, F. L., "Insurance Science and Economics". (Note Chapter 9 on application of averages.)

Jerome, Harry, "Statistical Method".

Jones, A. L., "Logic Inductive and Deductive, An Introduction to Scientific Method".

Jones, D. C., "A First Course in Statistics".

Jordan, D. F., "Practical Business Forecasting".

Karsten, Karl G., "Charts and Graphs".

Kelly, Truman L., "Statistical Method".

Kent, F. C., "Elements of Statistics".

Keynes, J. M., "A Treatise on Probability".

King, W. I., "Elements of Statistical Method".

King, W. I., "Index Numbers Elucidated".

Koren, John (Editor), "The History of Statistics; Their Development and Progress in Many Countries".

Kuznets, Simon S., "Secular Movements in Production and Prices" and "Seasonal Variations in Industry and Trade".

Lacroix and Ragot, "A Graphic Table Combining Logarithms and Antilogarithms".

Lipka, Joseph, "Graphical and Mechanical Computation".

Lovitt, W. V. and Holtzclaw, H. F., "Statistics".

Macaulay, F. R., "The Smoothing of Time Series".

Marshall, W. C., "Graphical Methods".

Mayo-Smith, Richmond, "Statistics and Sociology".

Meitzen, August, "History, Theory and Technique of Statistics".

Merriman, M., "The Method of Least Squares".

Miner, J. R. "Tables.....for use in Partial Correlation and Trigonometry".

Mitchell, Wesley C., Bulletin 284, Bureau of Labor Statistics, "Index Numbers of Wholesale Prices in the United States and Foreign Countries, 1921" (Revision of Bulletin 173).

Mitchell, Wesley C., "Business Cycles--The Problem and Its Setting".

Moore, Henry L., "Forecasting the Yield and Price of Cotton"; "Economic Cycles: Their Law and Cause", and "Generating Economic Cycles".

Mudgett, Bruce D., "Statistical Tables and Graphs".

Newsholme, Arthur, "The Elements of Vital Statistics".

Passano, L. M., "Calculus and Graphs".

Pearl, Raymond, "Medical Biometry and Statistics".

Pearson, Karl, "Tables for Statisticians and Biometricians".

Pearson, Karl, "The Grammar of Science, Physical".

Persons, Foster, and Hettinger, "Problem of Business Forecasting".

Persons, W. M., "The Construction of Index Numbers".

Riegel, Robert, "Elements of Business Statistics".

Riggleman, J. R., and Frisbee, I. N., "Business Statistics".

Reinhardt, J. M. and Davies, G. R., "Principles and Methods of Sociology".

Rietz, H. L., (editor), "Handbook of Mathematical Statistics".

Rietz, H. L., monograph on "Mathematical Statistics".

Rugg, H. O., "Statistical Methods Applied to Education".

Schultz, Henry, "Statistical Laws of Supply and Demand".

Secrist, Horace, "An Introduction to Statistical Methods".

Secrist, Horace, "Readings and Problems in Statistical Methods".

Slichter, C.S., "Elementary Mathematical Analysis" (Various editions).

Smith, B. B., "The Use of Punched Card Tabulating Equipment in Multiple Correlation Problems".

Smith, James G., "Elementary Statistics".

Snider, Joseph L., "Business Statistics".

Snyder, C., "Business Cycles and Business Measurements".

Social science research council. Advisory committee on social and economic research in agriculture, "Research Method and Procedure in Agricultural Economics" (2 volumes).

Sutcliffe, W. G., "Elementary Statistical Methods".

Thorndike, E. L., "An Introduction to the Theory of Mental and Social Measurements" (Various editions).

Thurstone, L. L., "Fundamentals of Statistics".

Tippett, L. H. C., "The Methods of Statistics" (An introduction mainly for workers in the biological sciences).

Walker, Helen M., "Studies in the History of Statistical Method".

Wallace, H. A. and Snedecor, G. W., "Correlation and Machine Calculation". Ames, Iowa, Iowa State College of Agriculture and Mechanic Arts.

Walsh, C. M., "The Measurement of General Exchange Value".

Weld, L. D., "Theory of Errors and Least Squares". (Introduction to measurement and properties of errors, with examples.)

West, Carl J., "Introduction to Mathematical Statistics".

Whipple, G. C., "Vital Statistics".

White, R. Clyde, "Social Statistics".

Whitaker, E.T. and Robinson, G., "The Calculus of Observations".

Wright, T. W. and Hayford, J. F., "Adjustment of Observations".

Young, B. F., "Statistics as Applied to Business".

Yule, G. U., "An Introduction to the Theory of Statistics". (Several editions are available. An excellent bibliography at the end of each chapter.)

Zizek, Franz, "Statistical Averages" (Translated by Warren M. Persons).

Logarithms

1. Purpose of logarithms.

From the point of view of pure mathematics, logarithms possess many interesting as well as useful properties; from the viewpoint of the computer, they provide the means of shortening certain types of computations and of performing others that would, for practical purposes, be virtually impossible otherwise.

By their use, multiplication and division are accomplished by addition and subtraction respectively; and involution and evolution are accomplished by simple multiplication and division respectively.

2. Review of Exponents.

It will be useful to recall that as a matter of definition, where a is any given number,,

$$a^2 = a.a; \qquad a^3 = a.a.a. \text{ ; and generally} \qquad (1)$$

$$a^n = a.a \ldots a \text{ (to n factors)}$$

From this it follows that;

$$a^2 . a^3 = (a . a).(a.a.a) = a^{2+3} = a^5, \text{ and generally,}$$

$$a^m.a^n = a.a \ldots .a \text{ (to m factors)}.a.a \ldots a \text{ (to n factors)}$$

$$a^m.a^n = a^{m+n} \qquad (2)$$

Consider the expression

$$\frac{a^3}{a^2} = \frac{a \cdot a \cdot a}{a \cdot a} = a^{3-2} = a$$

or more generally

$$\frac{a^m}{a^n} = \frac{a \cdot a \ldots a \cdot a \text{(to m factors)}}{a \cdot a \ldots a \text{ (to n factors)}} = a^{m-n} \qquad (3)$$

If $n > m$, then $m - n < 1$, corresponding to $a^2 \div a^3 = \frac{a \cdot a}{a \cdot a \cdot a} = \frac{1}{a} = a^{2-3} = a^{-1}$

or generalizing $a^{-n} = \frac{1}{a^n}$ (4)

If $m = n$, then $m - n = o$, corresponding to

$$\frac{a^m}{a^n} = 1 = a^m \cdot \bar{a}^m = a^o = 1 \qquad (5)$$

Consider $(a^2)^3 = a^2 \cdot a^2 \cdot a^2 = (a.a).(a.a).(a.a) = a^6$ and

generally $(a^n)^m = a^{mn}$ (6)

Consider $a^{\frac{3}{4}} \cdot a^{\frac{3}{4}} \cdot a^{\frac{3}{4}} \cdot a^{\frac{3}{4}} = (a^{\frac{3}{4}})^4 = a^{\frac{12}{4}} = a^3$, by Eq. (6)

Hence, $\sqrt[4]{a^3} = a^{\frac{3}{4}}$ since $(a^{\frac{3}{4}})^4 = a^3$

and generally $a^{\frac{m}{n}} = \sqrt[n]{a^m}$ (7)

Expressed in words,

In a fractional exponent, the denominator denotes the root that is to be extracted of the quantity secured by raising the base to the power indicated by the numerator of the fraction.

3. Definition of logarithm and antilogarithm.

Logarithms are merely exponents. For example, since $8 = 2^3$, it may be said that 3 is the logarithm of 8 to the base 2, or, more briefly, $\log_2 8 = 3$.

Generally, if $y = a^m$ (8)

$\log_a y = m$

This expression is read, "The logarithm of y to the base a is m." This relation is very important.

The antilogarithm is the number corresponding to a given logarithm. In eq.8, y is the antilogarithm of m.

4. Performing operations with logarithms.

Consider the table

$1 = 2^0$	$64 = 2^6$
$2 = 2^1$	$128 = 2^7$
$4 = 2^2$	$256 = 2^8$
$8 = 2^3$	$512 = 2^9$
$16 = 2^4$	$1024 = 2^{10}$
$32 = 2^5$	$2048 = 2^{11}$

If it is desired to multiply 32 by 64, it is known from the study of exponents (Eq. 2) and by reference to the above table that

$$32 \cdot 64 = 2^5 \cdot 2^6 = 2^{5+6} = 2^{11}$$

The latter is seen by the table to be 2048.

Note that $5 = \log_2 32$

$$6 = \log_2 64$$

$$11 = \log_2 2048 = \log_2 32 + \log_2 64 = \log_2 (32 \cdot 64)$$

It is evident that the process of multiplication is replaced by that of addition. That is, to multiply two or more numbers, it is only necessary to add their logarithms, assuming, of course, that tables are readily available, giving the relation of numbers to their

logarithms.

Consider also the following examples of division, involution, and evolution:

Division:

$$1024 \div 16 = 2^{10} \div 2^4 = 2^{10-4} = 2^6 = 64$$

$$\log_2 1024 = 10$$

$$\log_2 16 = 4$$

$$\log (1024 \div 16) = 10 - 4 = 6$$

Involution:

$$8^3 = (2^3)^3 = 2^{3 \cdot 3} = 2^9 = 512$$

$$\log_2 8 = 3$$

$$\underline{\qquad\qquad x\ 3}$$

$$\log_2 512 = 9$$

$$\log_2 8^3 = 3 \log 8 = 3 \cdot 3 = 9$$

Evolution:

$$\sqrt[5]{1024} = (2^{10})^{1/5} = 2^{10/5} = 2^2 = 4$$

$$\log_2 1024 = 10$$

$$10 - 5 = 2$$

$$\log \sqrt[5]{1024} = \tfrac{1}{5} \log 1024 = \tfrac{1}{5} \cdot 10 = 2$$

These simple examples illustrate the following relations, which follow from equations (1) to (8) inclusive.

Multiplication:

$$\log_a (m.n) = \log_a m + \log_a n \qquad (9)$$

Division:

$$\log_a \left(\frac{m}{n}\right) = \log_a m - \log_a n \qquad (10)$$

Involution:

$$\log_a m^n = n \log_a m \qquad (11)$$

Evolution:

$$\log_a \sqrt[n]{m} = \frac{1}{n} \log_a m \qquad (12)$$

5. Choice of base for a system of logarithms:

Any real number may be used as a base for a system of logarithms. In certain types of work, the Napierian, or Natural system of logarithms is used, in which the number e = 2.71828 ... is the base. For most routine work, however, the Briggs system of logarithms is used, in which the number 10 is the base. The choice of 10 as the base is purely a matter of practical convenience; the reason being that our number system is a decimal system (i.e. "deci" means 10).

Consider the relations:

$\log 1 = 0.000$	for $1 = 10^{0.000}$
$\log 2 = 0.3010$	" $2 = 10.^{0.3010}$
$\log 3 = 0.4771$	" $3 = 10^{0.4771}$
$\log 4 = 0.6021$	" $4 = 10^{0.6021}$
$\log 5 = 0.6990$	" $5 = 10^{0.6990}$
$\log 10 = 1.0000$	" $10 = 10^{1.0000}$
$\log 50 = 1.6990$	" $50 = 10^{1.6990}$
$\log 100 = 2.0000$	"$100 = 10^{2.0000}$
$\log 500 = 2.6990$	"$500 = 10^{2.6990}$
$\log 1000 = 3.0000$	"$1000 = 10^{3.0000}$

A graph of this relation is as follows:

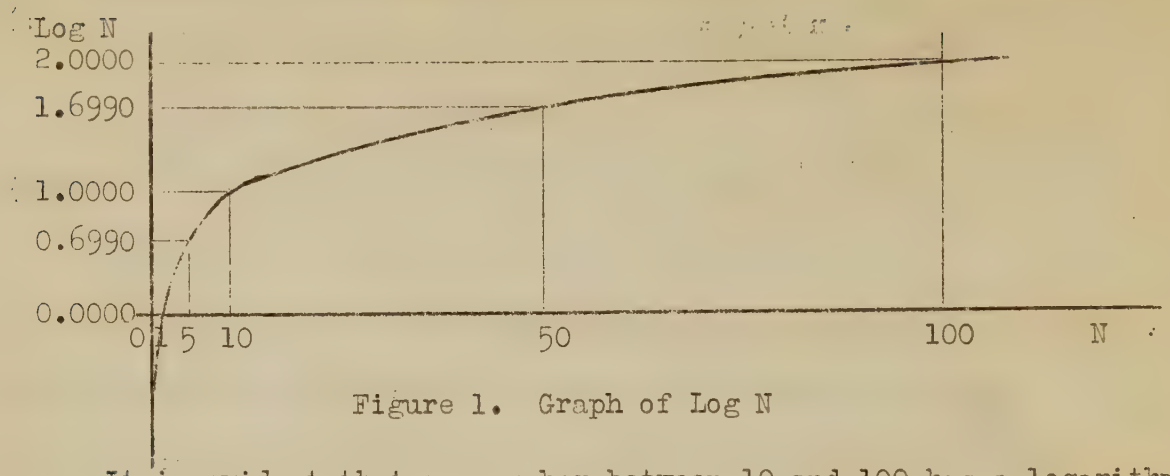


Figure 1. Graph of Log N

It is evident that any number between 10 and 100 has a logarithm between 1 and 2, that any number between 100 and 1000 has a logarithm between 2 and 3, and so on. Such logarithms may not, and usually are not, rational fractional powers of ten, but they may always be expressed to any desired degree of approximation.

6. Characteristic and Mantissa.

It will be noted that logarithms to the base 10 may be divided into two parts by the decimal point. The part to the left is called the characteristic, and the other is called the mantissa. The mantissa is dependent solely upon the sequence of digits in the number whose logarithm is sought. For example: the logs of 5, 50, and 500 each have the same mantissa, namely .6990. They differ only in the characteristic, because multiplication by 10 merely changes the logarithm by unity.

Mantissae are therefore tabled, and are located solely by the sequence of digits.

The characteristic is easily determined. The characteristic of any number between 100 and 1000 is 2, because such a number can be represented by 10 raised to 2 and a fractional power, the fraction

being the mantissa.

There are two simple rules for determining the characteristic of any positive number:

1. The characteristic of any number greater than 1 is positive; numerically it is one less than the number of places to the left of the decimal point. e. g. the characteristic of the log of 210 is 2, of 3564 is 3, etc.

2. The characteristic of a positive number less than one is negative; numerically it is one less than the number of zeros between the decimal point and the first significant figure e.g. The characteristic of the log of .123 is —1, of .00383 is —3, etc.

The student should verify these rules.

Exercise:

Determine the characteristics of the logs of the following numbers:

1. 5832	4. 82.46
2. 58.32	5. 138.23
3. .005832	6. .000005

7. Logarithm tables.

Logarithm tables are available to a varying number of places. Thus a four place table gives mantissae to 4 decimal places, corresponding to the sequences of three digits from 100 to 999, the first two digits being entered down the left hand side of the table, and the third across the top. Five place tables give mantissae to 5 decimal places, for sequences of digits from 1000 to 9999, the first three digits being entered down the left side of the table and the fourth across the top; and so on.

For more accurate work, five, six, eight, ten, or even twenty place tables may be required.

Exercise: Find the mantissae of the following numbers:

7. 383	10. 273
8. 839	11. 999
9. 111	12. 839

8. Interpolation.

Although a four place table gives mantissae for only three digits, it is possible to determine from such a table with fair accuracy the logarithm of a four digit number. Thus, if the log of 3458 is desired, the procedure is to find the mantissae of the logarithms of .3450 and 3460 (which are the same as for 345 and 346) and to calculate the mantissa of the log of 3458 from these two on the assumption that the change in the mantissa is proportional to the change in the argument. The figure illustrates this assumption.

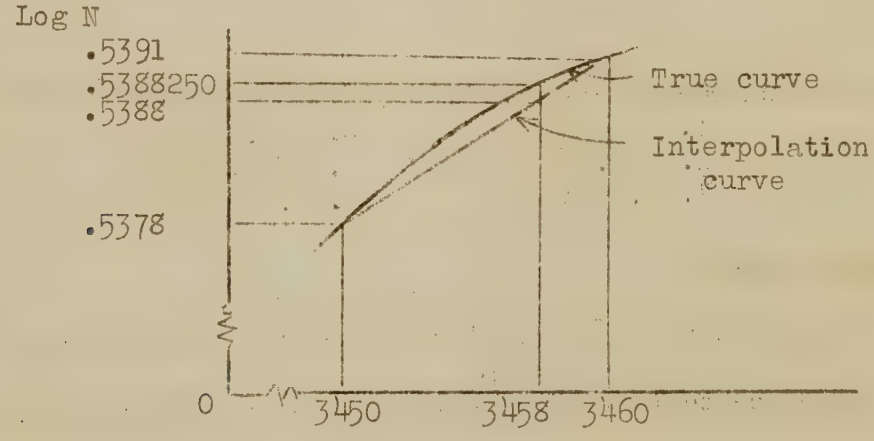


Figure 2. Interpolation

A convenient form for making this calculation is as follows:

Find the log of 34.58

Solution: log 3450 = 3.5378

3458 = 3.5388

3460 = 3.5391

Tabular difference = 3.5391 _ 3.5378 = .0013

.8 x .0013 = .0010

log 3458 = 3.5378 + .0010 = 3.5388

In Fig. 2 it will be observed that the true relation between n and log n is a curved line, whereas the method of interpolation just proposed assumes the relation is a straight line. There is a small error in such calculations, but it is usually not great enough to show up in the fourth place of the logarithm. As a rule however, not more than one extra digit should be handled in this manner from the usual table.

Exercises:

Find the logarithms of the following numbers:

13.	8324.	16.	.01001
14.	96.83	17.	.003861
15.	102,3	18.	99.99

9. Inverse interpolation.

Similarly when the argument corresponding to a given logarithm is desired, it is necessary to use inverse interpolation. Suppose it is desired to find the argument whose logarithm is 8753. A four place table gives a mantissa 8751 (corresponding to 750) and 8756 (corresponding to 751).

Since a change of 5 in the mantissa is accompanied by a change of 1 in the argument, it is assumed that a change of 2 in the log will be accompanied by a proportional change in the argument.

A convenient form of calculation is as follows:

Antilog $\left[\begin{array}{l|l} & 8751 = \\ 2 & \\ 5 & 8753 = \\ & 8756 \end{array}\right.$ $\begin{array}{l} 750 \\ 750 + (2/5 \times 1) = 750.4 \\ 751. \end{array}$

$$\frac{8753 - 8751}{8756 - 8751} = 2/5 = .4$$

Exercise:

Find the numbers corresponding to the following logs:

19.	3.3684	22.	.1528
20.	5.2691	23.	1.8205
21.	.8304	24.	1.9240

10. Conventions in using negative characteristics.

Since a number between zero and one has a negative characteristic and a positive mantissa, some convention is necessary for simplicity in writing. Thus the number .351 has the characteristic -1, and the mantissa .5453. This is sometimes written $\bar{1}.5453$. but more conveniently 9.5453 -10, both of which have a numerical value of − .4547. Either of these three forms may be used in calculations, the second, i.e. 9.5433 − 10, usually being most convenient.

Example 1. Multiply .350 by .0480

log .350 = 9.5441 − 10

log .0480 = 8.6812 − 10

18.2253 − 20 = 8.2253 − 10

Antilog 8.2292 − 10 = .0168 Ans.

Example 2. Divide .0480 by .350

log .0480 = 18.6812 − 20

log .350 = 9.5441 − 10

9.1371 − 10

Antilog 9.1371 − 10 = .1371

Exercises:

By means of logarithms to the base 10 perform the the following operations, and show your work:

1. $.0235 \times 3.14$
2. $.273 \times 12.0$
3. $\sqrt{.2908 \times .03054 \times 5634}$
4. $\sqrt[5]{(32)^3}$
5. $83.95 \div .42.83$
6. $\frac{32.54 \times 27.82 \times 12.34}{54.32 \times 82.27 \times 34.12}$
7. $\sqrt[8]{2.49 \times 65.74}$
8. $323 \times .323$
9. $\frac{86 \times 3845}{8^4}$
10. $5693 \div .8494$

No.	0	1	2	3	4	5	6	7	8	9
10	0000	0043	0086	0128	0170	0212	0253	0294	0334	0374
11	0414	0453	0492	0531	0569	0607	0645	0682	0719	0755
12	0792	0828	0864	0899	0934	0969	1004	1038	1072	1106
13	1139	1173	1206	1239	1271	1303	1335	1367	1399	1430
14	1461	1492	1523	1553	1584	1614	1644	1673	1703	1732
15	1761	1790	1818	1847	1875	1903	1931	1959	1987	2014
16	2041	2068	2095	2122	2148	2175	2201	2227	2253	2279
17	2304	2330	2355	2380	2405	2430	2455	2480	2504	2529
18	2553	2577	2601	2625	2648	2672	2695	2718	2742	2765
19	2788	2810	2833	2856	2878	2900	2923	2945	2967	2989
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784
24	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962
25	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133
26	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298
27	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609
29	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757
30	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900
41	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038
43	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172
44	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302
34	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428
35	5441	5453	5465	5478	5490	5502	5514	5527	5539	5551
36	5563	5575	5587	5599	5611	5623	5635	5647	5658	5670
37	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786
38	5798	5809	5821	5832	5843	5855	5866	5877	5888	5899
39	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010
40	6021	6031	6042	6053	6064	6075	6085	6096	6107	6117
41	6128	6138	6149	6160	6170	6180	6191	6201	6212	6222
42	6232	6243	6253	6263	6274	6284	6294	6304	6314	6325
43	6335	6345	6355	6365	6375	6385	6395	6405	6415	6425
44	6435	6444	6454	6464	6474	6484	6493	6503	6513	6522
45	6532	6542	6551	6561	6571	6580	6590	6599	6609	6618
46	6628	6637	6646	6656	6665	6675	6684	6693	6702	6712
47	6721	6730	6739	6749	6758	6767	6776	6785	6794	6803
48	6812	6821	6830	6839	6848	6857	6866	6875	6884	6893
49	6902	6911	6920	6928	6937	6946	6955	6964	6972	6981
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067
51	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152
52	7160	7168	7177	7185	7193	7202	7210	7218	7226	7235
53	7243	7251	7259	7267	7275	7284	7292	7300	7308	7316
53	7324	7332	7340	7348	7356	7364	7372	7380	7388	7396
No.	0	1	2	3	4	5	6	7	8	9

No.	0	1	2	3	4	5	6	7	8	9
55	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474
56	7482	7490	7497	7505	7513	7520	7528	7536	7543	7551
57	7559	7566	7574	7582	7589	7597	7604	7612	7619	7627
58	7634	7642	7649	7657	7664	7672	7679	7686	7694	7701
59	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774
60	7782	7789	7796	7803	7810	7818	7825	7832	7839	7846
61	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917
62	7924	7931	7938	7945	7952	7959	7966	7973	7980	7987
63	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055
64	8062	8069	8075	8082	8089	8096	8102	8109	8116	8122
65	8129	8136	8142	8149	8156	8162	8169	8176	8182	8189
66	8195	8202	8209	8215	8222	8228	8235	8241	8248	8254
67	8261	8267	8274	8280	8287	8293	8299	8306	8312	8319
68	8325	8331	8338	8344	8351	8357	8363	8370	8376	8382
69	8388	8395	8401	8407	8414	8420	8426	8432	8439	8445
70	8451	8457	8463	8470	8476	8482	8488	8494	8500	8506
71	8513	8519	8525	8531	8537	8543	8549	8555	8561	8567
72	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627
73	8633	8639	8645	8651	8657	8663	8669	8675	8681	8686
74	8692	8698	8704	8710	8716	8722	8727	8733	8739	8745
75	8751	8756	8762	8768	8774	8779	8785	8791	8797	8802
76	8808	8814	8820	8825	8831	8837	8842	8848	8854	8859
77	8865	8871	8876	8882	8887	8893	8899	8904	8910	8915
78	8921	8927	8932	8938	8943	8949	8954	8960	8965	8971
79	8976	8982	8987	8993	8998	9004	9009	9015	9020	9025
80	9031	9036	9042	9047	9053	9058	9063	9069	9074	9079
81	9085	9090	9096	9101	9106	9112	9117	9122	9128	9133
82	9138	9143	9149	9154	9159	9165	9170	9175	9180	9186
83	9191	9196	9201	9206	9212	9217	9222	9227	9232	9238
84	9243	9248	9253	9258	9263	9269	9274	9279	9284	9289
85	9294	9299	9304	9309	9315	9320	9325	9330	9335	9340
86	9345	9350	9355	9360	9365	9370	9375	9380	9385	9390
87	9395	9400	9405	9410	9415	9420	9425	9430	9435	9440
88	9445	9450	9455	9460	9465	9469	9474	9479	9484	9489
89	9494	9499	9504	9509	9513	9518	9523	9528	9533	9538
90	9542	9547	9552	9557	9562	9566	9571	9576	9581	9586
91	9590	9595	9600	9605	9609	9614	9619	9624	9628	9633
92	9638	9643	9647	9652	9657	9661	9666	9671	9675	9680
93	9685	9689	9694	9699	9703	9708	9713	9717	9722	9727
94	9731	9736	9741	9745	9750	9754	9759	9763	9768	9773
95	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818
96	9823	9827	9832	9836	9841	9845	9850	9854	9859	9863
97	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908
98	9912	9917	9921	9926	9930	9934	9939	9943	9948	9952
99	9956	9961	9965	9969	9974	9978	9983	9987	9991	9996
No.	0	1	2	3	4	5	6	7	8	9